Toxic expanded GGGGCC repeat transcription is mediated by the PAF1 complex in *C9orf72*-associated FTD.

Lindsey D. Goodman[1], **Mercedes Prudencio**[3], **Nicholas J. Kramer**[4], **Luis F. Martinez-Ramirez**[2], **Ananth R. Srinivasan**[2], **Matthews Lan**[2], **Michael J. Parisi**[2], **Yongqing Zhu**[2], **Jeannie Chew**[3], **Casey N. Cook**[3], **Amit Berson**[2], **Aaron D. Gitler**[5], **Leonard Petrucelli**[3], and **Nancy M. Bonini**[1,2,*]

[1]Neuroscience Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

[2]Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA

[3]Department of Neuroscience, Mayo Clinic, Jacksonville, FL 32224, USA

[4]Neuroscience Graduate Program, Stanford University School of Medicine, Stanford, CA 94305, USA

[5]Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA

Abstract

An expanded (G4C2)30+ repeat within *C9orf72* is the most prominent mutation in familial FTD and ALS. Through an unbiased, large-scale screen in (G4C2)49-expressing *Drosophila* we identify the CDC73/PAF1 complex (PAF1C), a transcriptional regulator of RNAPII, as a suppressor of G4C2-associated toxicity. Depletion of PAF1C reduces RNA and GR-dipeptide production from (G4C2)30+ transgenes. Interestingly, dPAF1C components, *dPaf1* and *dLeo1* appear selective for transcription of long, toxic repeat expansions, but not shorter, non-toxic expansions. In yeast, scPAF1C components regulate expression of both sense and anti-sense repeats. PAF1C is upregulated upon expression of (G4C2)30+ in flies and mice. hPaf1 is also upregulated in *C9+*-derived cells and its heterodimer partner, hLeo1, binds *C9+* repeat chromatin. In *C9+* FTD, *hPAF1* and *hLEO1* are upregulated and their expression positively correlates with expression of repeat-containing *C9orf72* transcripts. These data indicate that PAF1C activity is an important factor for transcription of the long, toxic repeat in *C9+* FTD.

INTRODUCTION

An intronic GGGGCC hexanucleotide (G4C2) expansion of >30 repeats found within *C9orf72* is the most prominent mutation in familial frontotemporal degeneration (FTD) and amyotrophic lateral sclerosis (ALS)[1,2]. These devastating neurodegenerative disorders represent a continuum of the same disease[3,4].

While the presence of the expanded G4C2 can confer toxicity through a number of aberrant pathways, amassing evidence supports contributions by gain-of-function mechanisms[5–7]. Accumulation of sense-strand G4C2- and antisense-strand G2C4-containing RNA results in RNA foci throughout the nervous system and may be toxic[8]. Further, these RNAs undergo repeat-associated non-AUG (RAN) translation to produce dipeptide-repeats (DPRs) – GA, GR, GP, PA, PR – which form potentially toxic aggregates[8]. Expressing (G4C2)30+ repeats, independent of changes to expression of the *C9orf72* gene product, causes toxicity in multiple model systems[5–7]. In *Drosophila*, expression of (G4C2)30+ transgenes results in RNA foci formation, DPR expression, and neurodegenerative effects[5,6].

Repeat-expansions, like (G4C2)n, have been defined in a number of neurodegenerative diseases. As these repeats tend to be GC-rich, they can form secondary structures (e.g. G-quadraplexes) and R-loops which are thought to diminish RNA polymerase II (RNAPII)-driven transcription[7,9–13]. Specialized transcriptional machinery may be required to promote the activity of RNAPII through such long repeat expansions[11,13]. Of note, DRB-sensitivity-inducing factor complex (DSIF) and the CDC73/PAF1 complex (PAF1C; PAF1 complex) are RNAPII regulators that may be involved since transcription of GC-rich DNA is sensitive to their loss[14–16].

DSIF and PAF1C are highly conserved and have non-redundant roles in activating RNAPII during elongation[17–20]. DSIF is composed of two proteins, Spt4 and Spt5 (SPT4H/5H in humans)[18]. SPT4H is implicated as a transcriptional regulator of expanded CAG and G4C2 repeats in disease[21–23]. PAF1C is composed of five proteins – Paf1, Leo1, CDC73, Ctr9, and Rtf1 – and it is unknown if hPAF1C has a role in repeat expansion diseases[19,20]. PAF1C downregulation can impact elongation rates of RNAPII, although only a subset of genes are known to require its function for expression[24–26]. In the nervous system, PAF1C is critical during development although it remains expressed in mature neurons[27–32].

Here, we identified PAF1C components as modifiers of *C9orf72*-associated disease in an unbiased, large scale RNAi-based screen in *Drosophila* expressing toxic (G4C2)49. Downregulation of PAF1C components disrupted transcription of the G4C2 RNA in both *Drosophila* and *S. cerevisiae* and resulted in reduced toxicity in *Drosophila*. Importantly, components *dPaf1* and *dLeo1* appear selective for transcription of long repeat transgenes, and not short repeats; depletion of other dPAF1C subunits and *dSpt4* impacted both long and short repeats. Furthermore, PAF1C is upregulated in response to expression of the expanded repeat in flies, mouse models, *C9+* patient-derived cells and *C9+* FTD tissue. Notably, expression of *hPAF1* and *hLEO1* positively correlate to expression of repeat-containing *C9orf72*-transcripts in patient tissue and hLeo1 is bound to *C9orf72* chromatin in *C9+* cells.

These data highlight hPAF1C as an important transcriptional regulator of expanded G4C2 within *C9orf72*.

RESULTS

RNAPII transcriptional complexes are enriched among suppressors of (G4C2)49-toxicity.

To define cellular mechanisms underlying (G4C2)30+ toxicity *in vivo*, a fly model was developed that expresses (G4C2)n repeats, utilizing the GAL4/UAS expression system (Fig. 1a). Quantification of the number of repeats inserted into individual transgenic lines highlighted four lines: control (G4C2)8, intermediate (G4C2)29, and two expanded (G4C2)49 lines with different insertion sites (Fig. 1b). These lines were analyzed for RNA level by northern blot to confirm similar expression (Fig. 1c). Expression of (G4C2)8, (G4C2)29, or (G4C2)49 transgenes in the fly optic system (Gmr-GAL4) revealed that (G4C2)49 expression caused degenerative effects: disruptions to the external ommatidial organization, red pigmentation, eye size, and internal retinal tissue loss (Fig. 1d). (G4C2)29 expression caused mild toxicity in 80% of animals: disruptions to the ommatidial organization and gaps within the internal retinal tissue.

An unbiased fly screen was developed that utilized the degenerative eye caused by (G4C2)49 expression with Gmr-GAL4 (Fig. 1e; Supplementary Fig. 1a). Using transgenic UAS-RNAi fly lines (see methods) we defined individual genes whose downregulation altered this toxicity with enhancement or suppression. Overall, 3,932 genes were tested: ~25% of the fly genome (Supplementary Data). 350 (8.9%) of these altered (G4C2)49-toxicity. Positive RNAi lines were then rigorously assessed to exclude those with unspecific effects: RNAi lines that altered normal eye morphology in control animals or altered expression of a control UAS-LacZ transgene (Supplementary Fig. 1b–c)[23]. Overall, 119 modifiers of (G4C2)49-toxicity were identified: 55 suppressors and 64 enhancers (Fig. 1f).

To define enriched processes, functions or components within these 119 modifiers, gene ontology (GO)-term analyses were performed (see methods). Enrichment scores (the degree to which a list of genes in a GO-term are represented within the modifier list) were plotted for GO-terms that were significantly represented within the panel of modifiers, p-value threshold of 10^−3 (Supplementary Fig. 1d; Supplementary Data). Among suppressors, there was strong enrichment for genes associated with RNAPII-driven transcription (Fig. 1g, **green**; Supplementary Data), including components of the dCDC73/PAF1 complex (dPAF1C; PAF1 complex) and dMediator complex[33]. Additional transcription regulators included *dELL* and *dEar* of the dSuper Elongation Complex (dSEC)[16] and *dSpt4* of the dDSIF complex. By contrast, enhancers of (G4C2)49-toxicity varied in GO-terms. Genes involved in RNA processing and splicing were identified, with components of the pre-catalytic spliceosome being the most enriched complex (Fig. 1h, **red**; Supplementary Data).

Overall, dPAF1C was the most enriched complex of transcriptional regulators that suppress (G4C2)49-toxicity.

PAF1C is selective for toxicity from a G4C2-encoding RNA.

Expanded G4C2 RNA can produce dipeptide-repeat proteins GR, GA, and GP in the fly[5–8]. Of these, (GR)30+ is strongly toxic[34]. As modifiers of (G4C2)49-toxicity may be acting downstream of GR-peptide production, we analyzed the 119 RNAi lines for their ability to modulate (GR)36-toxicity (Fig. 2a). GR models express a GR-peptide from a non-G4C2 RNA transcript, thus avoiding potential toxicity caused by a G4C2 repeat-bearing RNA[34]. As the PAF1 and DSIF complexes may be of particular interest in repeat-associated disease, dPAF1C and *dSpt4* RNAi were rigorously examined for modification of (GR)n-toxicity in flies and yeast (Supplementary Fig. 2a–c; for *scSpt4*[23]). Results showed no consistent or significant effects indicating that these modifiers have a minimal effect on (GR)n-associated toxicity.

In *Drosophila*, 48 of the 119 modifiers (40.3%) did not alter (GR)36-toxicity in the same manner that they altered G4C2-repeat toxicity, suggesting that they modulate expression of the RNA or RNA-derived toxicity (Fig. 2b; Supplementary Data). GO-term analysis of these modifiers revealed that suppression by dPAF1C components was selective to the (G4C2)49 model (Fig. 2c). In contrast, the dMediator complex and *dELL* and *dEar* of dSEC similarly suppressed both (G4C2)49- and (GR)36-toxicity in the external eye.

Overall, data suggested that dPAF1C suppression of (G4C2)49-toxicity was the result of effects upstream of toxic GR-peptide production.

dPAF1C is selective for the G4C2 expansion.

Biological connections between (G4C2)30+ expression and TAR DNA binding protein 43 (TDP43) pathology have been reported in *C9+* FTD/ALS[5,8]. Thus, we examined the (G4C2)49 modifiers in a TDP43 model to define G4C2-unique pathways (Fig. 2a). Expression of human TDP43 in the fly eye causes toxicity[35]. Rigorous assessment of dPAF1C and *dSpt4* RNAi showed that they had no significant effect on TDP43-toxicity (Supplementary Fig. 3a–b). 56 of the 119 (G4C2)49 modifiers (47.1%) were unique to (G4C2)49 (Fig. 2b; Supplementary Data). GO-term analysis indicated that dPAF1C was again highly enriched, underscoring specificity of this complex to the expanded G4C2 (Fig. 2d). By contrast, other RNAPII-regulators, such as the dMediator complex and *dELL* and *dEar* of dSEC, similarly suppressed (G4C2)49- or TDP43-toxicity in the external eye (also[35]). Furthermore, dPAF1C RNAi did not alter TDP43 protein levels (Supplementary Fig. 3c).

Taken together, these data highlight dPAF1C selectivity towards (G4C2)49-associated toxicity.

dPAF1C downregulation suppresses (G4C2)49-induced toxicity in fly.

We further examined if dPAF1C downregulation could impact (G4C2)49-toxicity in multiple fly tissues. First, we confirmed that the RNAi lines targeting dPAF1C components – *dPaf1 (ATMS)*, *dLeo1 (ATU)*, *dCDC73 (HYX)*, *dCtr9* and *dRtf1*–resulted in significant downregulation of the target gene, while extending reagents to include an independent set of dPAF1C RNAi lines (Supplementary Fig. 4).

Consistent with effects in the external eye, co-expression of dPAF1C RNAi with (G4C2)49 rescued internal tissue loss (Fig. 3a–b). *dPaf1*, *dLeo1*, and *dRtf1* RNAi caused ~80% recovery and *dCDC73* and *dCtr9* RNAi caused ~100% recovery of tissue depth: control =13.4±6.2μm, *dPaf1* =44.4±4.7μm, *dLeo1* =47.2±10.8μm, *dCDC73* =63±6.1μm, *dCtr9* =69.1±18.0μm, *dRtf1* =41.8±12.7μm. This indicates that dPAF1C depletion is markedly effective in mitigating (G4C2)49-toxicity in the eye. Downregulating dPAF1C components did not affect the external or internal eye on their own (Fig. 3c–d). The second set of dPAf1C RNAi lines also suppressed (G4C2)49-toxicity, while not affecting the normal fly eye, affirming that suppression is indeed the result of targeting dPAF1C (Supplementary Fig. 5a–b). To confirm the effect on G4C2-induced toxicity, an independent (G4C2)30+ disease fly model was examined for interactions with dPAF1C. This model contains 114bp of the intronic sequence found upstream of G4C2 repeat in patients (termed "leader sequence"; LDS). Again, dPAF1C RNAi suppressed toxicity caused by LDS-(G4C2)44[GR-GFP] expression within the fly optic system (Supplementary Fig. 5c).

To examine effects of downregulating dPAF1C components in the nervous system, we expressed the (G4C2)49 transgene using a conditional neuronal GAL4 driver, ElavGS. First, we assessed climbing ability of animals with adult-onset expression. At 14d, only 41% of (G4C2)49-expressing animals were able to climb 4cm up the wall of a vial within 20s (Fig. 3e). dPAF1C RNAi significantly suppressed this effect, causing climbing abilities to be maintained at 92–96%. Suppression was again seen when using the second set of dPAF1C RNAi lines (Supplementary Fig. 5d).

Next, we analyzed animals for age-dependent vacuole formation in the fly brain, an indicator of neurodegeneration. (G4C2)49 expression in the adult-fly nervous system caused large vacuoles throughout the brain at 28d. *dCDC73*, *dCtr9*, and *dRtf1* RNAi significantly reduced both vacuole size and number: mean degeneration score (3.4±0.5) was reduced to 2.10±0.6 (*dCDC73* RNAi), 2.38±0.5 (*dCtr9* RNAi), and 1.78±0.8 (*dRtf1* RNAi), with the normal brain being 1.1±0.4 (Fig. 3f; Supplementary Fig. 6). Further, the strong suppressor *dCDC73* could increase adult survival of (G4C2)49-expressing animals, with 50% survival extended by 24h and end point extended 72h (Fig. 3g). RNAi targeting other dPAF1C components showed reduced lifespan in control flies, making it inaccurate to evaluate their activity as suppressors of (G4C2)49 lifespan (Supplementary Fig. 7a). We further rigorously ruled out any possible genetic background effects of the RNAi lines within the context of (G4C2)49-expression (Supplementary Fig. 8).

Together, data support that dPAF1C plays an important role in (G4C2)49-induced toxicity in disease-relevant tissues in the fly.

PAF1C mediates expression of expanded G4C2 in fly and yeast.

Given that PAF1C regulates RNAPII-elongation and is important for transcription of GC-rich DNA[15,19,20], we hypothesized that PAF1C may be important for successful transcription of long G4C2 repeats.

We first determined whether downregulation of dPAF1C components – *dPaf1*, *dLeo1*, *dCDC73*, *dCtr9*, and *dRtf1* – affected the amount of RNA produced from expanded

(G4C2)49 transgenes in the fly nervous system, using *dSpt4* RNAi as a positive control[23] (Fig. 4a, **red**). dPAF1C RNAi caused significant reductions in (G4C2)49 RNA: *dPaf1* =41±8% reduction, *dLeo1* =54±4% reduction, *dCDC73* =57±5% reduction, *dCtr9* =62±6% reduction, *dRtf1* =62±5% reduction. These values were similar to *dSpt4* RNAi which caused a 52±5% reduction (also[23]). This effect was consistent in a second (G4C2)49-fly model (Supplementary Fig. 9a).

We evaluated the specificity of dPAF1C alterations on expanded (G4C2)49-RNA, by assessing expression from a nontoxic (G4C2)8 and an intermediate (G4C2)29 repeat. Surprisingly, *dPaf1* and *dLeo1* were selective for expanded (G4C2)49 as these RNAi did not significantly reduce expression from the shorter, nontoxic repeat transgenes (Fig. 4a, **black and green**). By contrast, *dSpt4*, *dCDC73*, *dCtr9*, and *dRtf1* RNAi caused significant reductions in expression from (G4C2)8 and (G4C2)29 of >30%, similar to effects on the (G4C2)49 transgene. Statistical comparisons between each (G4C2)n transcript levels with individual RNAi (Fig. 4a, **top**), confirmed that *dPaf1* and *dLeo1* RNAi show no significant difference between the expression level of (G4C2)8 and (G4C2)29 but caused a statistically significant drop between (G4C2)<30 and (G4C2)49 RNA levels. By contrast, *dSpt4* RNAi and other dPAF1C RNAi had statistically similar decreases in RNA levels among short and expanded G4C2 repeat lengths. These findings suggest that, of the dPAF1C components, *dPaf1* and *dLeo1* appear of special importance for expression of longer repeats in the fly. We further confirmed that the reduced expression was specific to the G4C2 construct by examining expression of an alternative disease transgene, TDP43. TDP43 transcript levels were not affected by dPAF1C RNAi in the fly brain, consistent with protein data (Fig. 4b; Supplementary Fig. 3c).

As (G4C2)30+ RNA can produce a toxic GR-dipeptide, we next asked if reduced expression of dPAF1C could alter GR production in the LDS-(G4C2)44[GR-GFP] fly model (Fig. 4c). The GFP tag in the GR reading frame allows for fluorescence imaging and quantitation of GR-levels in the eye with Gmr-GAL4. Congruent with RNA data, dPAF1C RNAi significantly downregulated GR-GFP: *dPaf1* =40±3% reduction, *dLeo1* =42±2% reduction, *dCDC73* =56±1% reduction, *dCtr9* =50±3% reduction, *dRtf1* =51±2% reduction.

These findings were then extended to yeast to assess if scPAF1C could regulate expression of expanded G4C2 in another model system. Transgenes were expressed from a galactose-inducible promoter. *scLeo1* (*leo1*Δ) or *scCDC73* (*cdc73*Δ) deletion had mild or no effect on expression from a control eYFP transgene (Fig. 4d, **black**). Deletion of other scPAF1C components had growth defects and/or significantly repressed expression of control eYFP so could not be assessed (data not shown). Interestingly, RNA levels from an expanded (G4C2)66 transgene were significantly decreased by *leo1*Δ or *cdc73*Δ: 40±14% and 38±2.5% reduction, respectively (Fig. 4d, **red**). Further, as the hexanucleotide repeat in *C9+* FTD/ALS patients can be transcribed bidirectionally[8], we determined if scPAF1C mutants could also regulate expression from an antisense-(G2C4)66 transgene. Notably, *leo1*Δ or *cdc73*Δ had an even stronger effect on expression from this transgene with *leo1*Δ showing a significantly stronger effect than *cdc73*Δ: 77±4% and 52±11% reduction, respectively (Fig. 4d, **orange**). These data confirmed that PAF1C downregulation impaired expression of

expanded G4C2 repeats in an independent model system, and extended data to expanded antisense-G2C4 repeat.

Effects of PAF1C RNAi could be the result of co-regulation of PAF1C components – *Paf1*, *Leo1*, *CDC73*, *Ctr9*, and *Rtf1* – with each other and/or *Spt4*. To examine this, we expressed *dPaf1*, *dLeo1*, or *dCDC73* RNAi in adult flies using the drug-inducible ubiquitous driver, DaGS. qPCR was used to determine changes in expression from dPAF1C components not targeted by the RNAi or *dSpt4* (Supplementary Fig. 7b). Data indicated that depleting these dPAF1C components does not alter expression of *dSpt4* or alternative dPAF1C components. Further, we confirmed that downregulating PAF1C in flies and in yeast did not alter general RNAPII transcription of endogenous housekeeping genes (Supplementary Fig. 9b–c)[19,20,24–26].

These data support that PAF1C mediates transcription of expanded G4C2 and G2C4. Further, data in the fly argues that *dPaf1* and *dLeo1* confer repeat-length specificity to the transcriptional machinery, promoting RNAPII-driven transcription of expanded (G4C2)30+ repeats.

PAF1C is upregulated in response to expanded G4C2.

Given the important role dPAF1C is playing in (G4C2)49-expressing animals, we considered that the complex may be dysregulated in *C9+* situations. Thus, we examined the expression of endogenous dPAF1C components – *dPaf1*, *dLeo1*, *dCDC73*, *dCtr9*, and *dRtf1* – in animals expressing (G4C2)49 in the adult fly nervous system (ElavGS, 16d). Surprisingly, all components were upregulated by qPCR (Fig. 5a): *dPaf1* =69±10% upregulation, *dLeo1* =58±13% upregulation, *dCDC73* =71±17% upregulation, *dCtr9* =49±12% upregulation, and *dRtf1* =64±5% upregulation. This effect did not occur in response to short (G4C2)8 repeat expression. Further, dPAF1C upregulation appeared selective to G4C2 as dPAF1C components were not upregulated in TDP43 expressing animals, but rather downregulated (Fig. 5b).

To determine if this same response to expression of the expanded G4C2 repeat occurred in mammals, protein levels of mLeo1 were measured in cortical tissue from (G4C2)2 or (G4C2)149 expressing mice (Fig. 5c). Transgenes were expressed using an AAV2/9 vector which predominantly transduces into neurons (see methods). At 3mo post-injection, no significant differences in mLeo1 were detected between cohorts. However, at 6mo a significant upregulation of 30±8% in mLeo1 was observed in (G4C2)149 animals compared to (G4C2)2 animals.

hPaf1 is upregulated in *C9+* patient-derived cells.

Our data from flies and mice indicated that PAF1C is upregulated in response to expression of an expanded G4C2 repeat in the brain. To better understand this effect in *C9+* disease, we extended our studies to patient-derived cells. Protein expression levels of four of the hPAF1C components – hPaf1, hLeo1, hCDC73, and hRtf1 – were assessed by western immunoblot in iPS cells (Fig. 6a). The mean expression from three *C9+* patient-derived cell lines was compared to three control cell lines. Notably, hPaf1 and hRtf1 were significantly upregulated by 46±14% and 54±14%, respectively, in *C9+* derived iPS cells.

hLeo1 binds *C9orf72* within the genome.

Thus far, data supported that the levels of select PAF1C components are modulated in response to expression of the repeat in flies, mice, and *C9+* patient-derived cells. Further, PAF1C is important for expression of expanded G4C2 in flies and yeast. Therefore, we considered that hPAF1C may be recruited and bound to *C9orf72* within the genome of patient-derived cells.

To assess this, chromatin immunoprecipitation (ChIP) studies were performed. Four, independent *C9+*-derived fibroblast lines (Supplementary Fig. 10a) were assayed, using a hLeo1 antibody (Fig. 6b). Importantly, Leo1 forms a heterodimer with Paf1 within PAF1C when bound to RNAPII[36,37]. After pull-down, the presence of the *C9orf72* gene or an *intergenic* (noncoding) sequence was assessed by qPCR and the mean enrichment was calculated from the four lines. Primers used to detect chromatin fragments from the *C9orf72* gene targeted the intronic region immediately 3' of the repeat expansion. Notably, there was a 4.9 ± 1.8-fold enrichment of *C9orf72*-sequence over the *intergenic*-sequence and IgG controls with hLeo1 ChIP, indicating that hLeo1 was bound to the *C9orf72* gene in patient-derived *C9+* cells. Overall, data support that hLeo1 binds the *C9orf72* gene and that it is bound through the repeat.

hPAF1 and *hLEO1* are upregulated in *C9+* FTD and FTD/ALS.

Our data from multiple model systems indicated that PAF1C is important for expression of the expanded G4C2 and that PAF1C is upregulated in response to expression of the repeat. To better understand the role of hPAF1C in human disease, we extended our studies to FTD/ALS patient tissue.

RNA was extracted from frontal cortex tissue of *C9+* patients (n=67), *C9-* patients (n=56), or healthy controls (n=27) and the level of *hPAF1* and *hLEO1* expression were assessed by qPCR (Fig. 6c; Supplementary Fig. 10b). In patients diagnosed with FTD (cortical disease), *hPAF1* and *hLEO1* were significantly upregulated only when the repeat expansion was present: *hPAF1* was upregulated 39% versus healthy controls and 28% versus *C9-* FTD; *hLEO1* was upregulated 23% versus healthy controls and 25% versus *C9-* FTD. Interestingly, in the frontal cortex of *C9+* ALS cases (motor neuron disease), neither *hPAF1* nor *hLEO1* showed upregulation, independent of the presence of the G4C2-repeat expansion. Patients showing symptoms of both FTD and ALS fell between FTD only and ALS only, with *C9+* FTD/ALS cases showing weaker upregulation of *hPAF1* and *hLEO1* versus healthy controls: 28% and 22%, respectively. In contrast to *hPAF1* and *hLEO1*, *hCDC73* did not show altered expression in *C9+* FTD (Supplementary Fig. 10c), supporting data in *Drosophila* and patient-derived cells that Paf1 and Leo1 may play a unique role over other components.

Upregulation of *hPAF1* and *hLEO1* in *C9+* FTD is congruent with data in flies, mice, and iPS cells. Interestingly, in cortical tissue this upregulation is unique to *C9+* FTD patients compared to *C9+* ALS patients (which lack cortical diagnosis).

Expression of *hPAF1* and *hLEO1* positively correlate with repeat-containing *C9orf72* transcripts.

As *Paf1* and *Leo1* activity were important for the expression of the expanded G4C2 repeat in flies and yeast, and hLeo1 is bound to *C9orf72* in *C9+* cells, we considered that *hPAF1* and *hLEO1* upregulation in *C9+* FTD may positively correlate with the expression of *C9orf72* transcripts containing the repeat.

To examine this, Spearman r correlations were performed between transcript levels of *hPAF1* and *hLEO1* versus *C9orf72* pre-mRNA (Fig. 6d). Results showed strong (r>0.3 to r>0.6) and significant (p-value <0.0005) positive correlations between transcripts in *C9+* FTD cases – for *hPAF1* vs *C9orf72*: r=0.74, 95% CI 0.5 to 0.9; for *hLEO1* vs *C9orf72*: r=0.63, 95% CI 0.3 to 0.8. No correlations were observed in *C9-* FTD cases or healthy controls supporting that this effect was in response to the presence of the repeat expansion. Further, no correlations between *hCDC73* transcripts and *C9orf72* pre-mRNA were observed, arguing that this effect was unique to *hPAF1* and *hLEO1* (Supplementary Fig. 10d). No correlations were observed in *C9+* ALS cases (Supplementary Fig. 10e).

Altogether, data from multiple models and FTD cortical tissue indicate that PAF1C regulates expression of the expanded G4C2 repeats within *C9orf72* in *C9+* situations.

DISCUSSION

An unbiased, RNAi-based screen in *Drosophila* covering ~4000 genes revealed the CDC73/PAF1 complex (PAF1C) as a *C9orf72*-disease modifier. Downregulation of dPAF1C components – *dPaf1*, *dLeo1*, *dCDC73*, *dCtr9*, and *dRtf1* – selectively suppressed (G4C2)49-toxicity in multiple fly tissues. Mechanistically, PAF1C suppression was associated with reduced RNA and GR-dipeptide production from expanded (G4C2)n transgenes. In particular, depletion of *dPaf1* and *dLeo1*, which form a heterodimer[36,37], in the fly nervous system selectively reduced expression of long, toxic (G4C2)49 repeats versus shorter, non-toxic repeats. Moreover, *scLeo1* and *scCDC73* deletion in yeast reduced RNA production from sense-(G4C2)66 and antisense-(G2C4)66 transgenes. Additional investigations into dPAF1C in *Drosophila* revealed that endogenous dPAF1C was upregulated upon expression of expanded (G4C2)30+ in neurons. This was not the result of toxicity or stress, as upregulation was selective to (G4C2)30+ versus TDP43. Interestingly, mLeo1 is also upregulated in the brain of mice expressing (G4C2)149, while hPaf1 and hRtf1 are upregulated in *C9+*-derived iPS cells. Further, we provide evidence that hPAF1C binds *C9orf72*, with ChIP for hLeo1 (of the Paf1/Leo1 heterodimer) in *C9+* cells. Using human post-mortem cortical tissue, we further found that *hPAF1* and *hLEO1* RNA levels were upregulated in *C9+* FTD patients. This upregulation positively correlated with expression of the repeat-containing *C9orf72* transcripts. Overall, these data from flies, yeast, mice, *C9+* patient-derived cells, and post-mortem patient tissue support a mechanistic link between PAF1C activity, expression of a G4C2-repeat, and *C9+* disease.

A total of 119 modifiers of (G4C2)49-toxicity were identified in our screen (**see** Fig 1). Remarkably, GO term analysis of the 55 suppressors showed significant enrichment for RNAPII transcriptional regulators – including dPAF1C, the dMediator complex, *dELL* and

dEar of the dSuper Elongation Complex (dSEC), and *dSpt4* of the dDSIF complex. Not all transcription-related genes tested in the screen modify (G4C2)49 toxicity, arguing that the complexes identified are unique (see Supplementary Data). Although we focused on dPAF1C, the other RNAPII regulators may also be important in disease[5–8]. Among the 64 enhancers of (G4C2)49-toxicity, RNA processing and splicing factors were prominent. Curiously, the majority of these enhancers similarly modulate (GR)36-toxicity (see Fig. 2). These data are consistent with reports of RNA dysregulation and splicing deficits in *C9+* FTD/ALS, while suggesting that disruptions in RNA metabolism may result from toxic GR[5–8].

RNAPII-driven transcription across GC-rich DNA is hypothesized to be problematic due to the propensity of the DNA/RNA to form secondary structures, such as R-loops and G-quadraplexes[7,9–13]. Given this, the activity of multiple elongation factors may be required for efficient transcription through expansions like (G4C2)30+[11,13]. *Spt4* was previously implicated as a transcriptional regulator of CAG and G4C2 repeat expansions[21–23]. Our data indicates that PAF1C regulates expression of G4C2 repeats in FTD/ALS (**see** Fig. 4). *dPaf1* and *dLeo1* of dPAF1C seem particularly important for RNAPII-transcription of expanded G4C2 as loss of these two components selectively reduced expression of a (G4C2)49 transgene in fly. While this could be the result of level of knockdown, it is compelling that Paf1 and Leo1 form a heterodimer that is important for PAF1C activation of elongating RNAPII[25,36–39]. Further, Paf1 is consistently upregulated in *C9+*-derived patient cells and tissue over other components and upregulation positively correlates to expression of repeat-containing transcripts in patient tissue (**see** Fig. 6c–d and Supplementary Fig. 10). Overall, these data argue that Paf1 and Leo1 are mechanistically special in *C9+* disease.

In contrast to *dPaf1/Leo1*, *dSpt4* showed similar effects on different G4C2 repeat lengths in flies (**see** Fig. 4). Differences in how Spt4 and Paf1/Leo1 loss impact expanded G4C2 expression may be due to their distinct roles during transcription[17–20]. Evidence suggests that Spt4 primarily acts during the transition of RNAPII from poised to elongation or during transcription termination[17,25,26]. In contrast, Paf1/Leo1 seems to act primarily during elongation[17,25,26,38,40]. Notably, DSIF (composed of Spt4 and Spt5) and PAF1C may interact. In yeast, scDSIF recruits scPAF1C via the scRtf1 subunit during the transition of RNAPII from initiation to elongation[41]. However, in higher organisms, including *Drosophila*, the dependence of PAF1C on DSIF for recruitment to elongating RNAPII is less clear. Rtf1 is less tightly associated with other PAF1C components while recent work shows that PAF1C recruitment can be DSIF-independent in mammals[42–46]. Further, Leo1 of PAF1C directly interacts with elongating RNAPII[37] and the *C9orf72* gene (**see** Fig. 6b). Together, these data suggest that PAF1C is playing a unique role in transcription elongation across the G4C2-repeat expansion.

FTD and ALS represent the extremes of a continuous disease spectrum. It remains unclear how one patient presents with FTD and another with ALS or combined FTD/ALS. Among many factors, unique genetic backgrounds of individuals may contribute to specific presentation[3,4]. For example, *ATXN2* and *TMEM106B* have been suggested as disease modifiers underlying the different diagnoses[3,4]. Curiously, we see selective upregulation of *hPAF1* and *hLEO1* in the frontal cortex of *C9+* FTD disease, but not *C9+* ALS (**see** Fig. 6c).

Importantly, the frontal cortex is thought to be a primary brain region resulting in FTD-associated symptoms[47,48]. Stage of disease progression could also contribute. Analysis of *hPAF1* and *hLEO1* expression in motor neurons from ALS patients will define if PAF1C plays a tissue-specific role. Based on current data, it is tempting to hypothesize that cortical modulation of *Paf1* or *Leo1* in response to the repeat may be among several mechanisms contributing to the spectrum of disease phenotypes in *C9orf72*-associated disorders.

PAF1C may be an attractive therapeutic candidate for *C9+* FTD. In yeast, *scPAF1C* has been reported to be non-essential[19,20]. In flies, *dLeo1* is not essential to get viable adults[49]. In mice, *mPAF1*$^{+/-}$ or *mLEO1*$^{+/-}$ heterozygosity yields no obvious abnormalities, although *mPAF1*$^{-/-}$ or *mLEO1*$^{-/-}$ null animals show pre-weaning lethality[50]. Other components of mPAF1C follow this same trend, with heterozygous mice showing no or few effects[50]. While *SUPT4H1* (human *Spt4*; non-essential in yeast) was previously proposed as a potential therapeutic target in *C9+* FTD/ALS[23], it may have more critical organismal functions than PAF1C in higher organisms: *SUPT4A*$^{+/-}$ (murine *Spt4*) heterozygous mice are viable but show a number of abnormalities and *SUPT4A*$^{-/-}$ null mice are embryonic lethal[50]. These investigations do not consider effects of PAF1C or Spt4 loss in a tissue- or age-specific manner, arguing that further analyses are required. Altogether, we hypothesize that specific components of PAF1C, like *Paf1* or *Leo1*, may represent a potential therapeutic target for *C9orf72*-associated disease.

This study presents the first evidence that PAF1C is an important player in *C9orf72*-associated disease, particularly in *C9+* FTD. Further investigations into PAF1C may define the mechanisms by which it becomes upregulated—potentially in a tissue-specific manner—and its impact in other neurodegenerative situations that result from aberrant expression of repeat expansions.

METHODS AND MATERIALS

Nomenclature

For clarity "d" for *Drosophila melanogaster*, "h" for *Homo sapiens*, "m" for *Mus musculus*, or "sc" for *Saccharomyces cerevisiae* was added in front of gene/protein symbols when discussing endogenous PAF1C components in individual species.

Patient sample consents and approvals

Participants or authorized family members provided written informed consent prior to information gathering. Post-consent, autopsies were performed postmortem. Protocols were approved by the Mayo Clinic Institutional Review Board and Ethics Committee.

Clinical, genetic and pathological assessments

Patients were diagnosed with FTD and/or ALS by trained neurologists after reviewing neurological and pathological information (Supplementary Fig. 10). Repeat-primed polymerase chain reaction was used to determine the presence/absence of a *C9orf72* repeat expansion[1].

Mouse model and approvals

All mouse procedures were performed in agreement with the National Institutes of Health Guide for Care and Use of Experimental Animals. Approved by the Mayo Clinic Institutional Animal Care and Use Committee. The G4C2 mouse model (C57BL/6J, male mice) were previously established[51].

Drosophila work and disease models

Stocks were maintained on standard cornmeal-molasses medium. Supplementary Table 2 and Supplementary Data detail fly lines used. For all experiments, multiple w^- and w^+ controls were analyzed in parallel and showed similar results. Supplementary Table 3 details shown controls.

Fly RNAi efficacy

All control and PAF1C RNAi lines defined in Supplementary Table 2. RNAi efficacy was determined using Da-GAL4 (larvae) or DaGS (adult animals), previously described[23,52]. For Paf1 RNAi: larvae were collected using 20% sucrose solution. For Leo1 RNAi: wandering 3rd instar larvae were collected.

Characterization of (G4C2)n fly models

All G4C2 fly models: Transgenes inserted into pUAST vectors and randomly inserted into w^{1118} fly genomes. Original G4C2 model detailed in Fig. 1a and previously established[23,53,54]. LDS-G4C2 model includes: 5' leader sequence (LDS; 114bp of sequence found upstream of the repeat in intron 1 of *C9orf72* in patients); 3'-GFP tag in the GR-reading frame. **Repeat-length determination:** Inserted transgenes were amplified from genomic DNA using construct-specific primers that flanked the repeat (Supplementary Table 4) by PCR: KAPA HiFi HotStart kit (Kappa #KK2501) with GC buffer, 1M Betaine, 5% DMSO. PCR products lengths were analyzed: 1.5% agarose/TAE gel, Bioanalyzer. Repeat number was calculated from PCR product length (subtracting 5' and 3' non-G4C2 flanking sequence). Control w^{1118} animals were included and showed no signal. **RNA expression for original UAS-G4C2 model:** UAS-G4C2 transgenes were expressed using HS-GAL4, 30min at 37°C. RNA was extracted 3hr post-HS and analyzed by northern blot, probing for the SV40 terminal sequence, previously described[55]. Control w^{1118} animals were included and showed no signal.

RNAi external fly eye screen

UAS-RNAi lines developed by the Transgenic RNAi Project (TRiP)[56,57] were purchased through Bloomington *Drosophila* Stock Center (BDSC). All available Valium 20 UAS-shRNA lines targeting unique genes were used for screening (detailed in Supplementary Data). When multiple Valium 20 fly lines were available targeting a single gene, one line was randomly selected for screening. Additional UAS-RNAi lines targeting *dPaf1* and *dLeo1* were obtained from Vienna *Drosophila* Resource Center (VDRC)[58]. **External eye screening and imaging:** UAS-RNAi males were crossed to recombinant females: UAS-(G4C2)49, Gmr-GAL4 (III) (25°C). Multiple controls were setup with every experiment to account for natural variability in (G4C2)49-toxicity, including: UAS-Luc RNAi (BDSC

#31603) and w^{1118}; UAS-DSRED. The external eye phenotype for >5 1–2d progeny was observed using a standard dissection microscope. Any changes to the ommatidial organization, eye size, pigmentation, and ability to eclose from pupae were noted. Resulting phenotype was categorized into one of six groups: suppressors, mild suppressors, no effect, mild enhancers, enhancers, and lethal enhancers (Supplementary Fig. 1a). Animals expressing RNAi that altered (G4C2)49-toxicity were imaged on a Leica Apo16 microscope[35,59–61]. Researchers were blinded to the RNAi targets during screening. Modifiers of (G4C2)49-toxicity were further assessed in Gmr-GAL4 > UAS-(GR)36[34] or UAS-TDP43[59,60] animals. Modifier crosses were repeated 3+ independent times to confirm reproducibility of results. **Control experiments defining unspecific RNAi:** To determine if RNAi caused toxicity in control scenarios: UAS-RNAi lines that enhanced (G4C2)49-toxicity were tested for effects in a line expressing short repeats, Gmr-GAL4 > (G4C2)6/8, and with Gmr-GAL4 only. UAS-RNAi lines were analyzed for effects on the GAL4/UAS system using β-galactosidase western immunoblots, described[23,35,54,61].

Gene ontology (GO) enrichment analysis

GO-term enrichment analyses were done using GOrilla software with a p-value threshold of 10^{-3}[62–64]. All genes included in the RNAi library, excluding unspecific RNAi, were set as the background. GO-terms with enrichment scores < 3.0 were excluded as these associated with umbrella categories. For final figures, redundant GO-terms were excluded using Revigo, prioritizing terms with high enrichment scores[64,65].

Internal eye and vacuole formation

Heads were paraffin embedded and sectioned at 8μm, described[23,35,61,66]. Quantification of internal eye: measured depth of the retinal tissue at a consistent level of the brain (when the optic chiasm, antennal lobe and ventrolateral protocerebrum were present); retinal tissue depth was measured at the point of the optic chiasm. Animals with collapsed eyes were excluded. Quantification of vacuole formation was done using a scoring schematic (Supplementary Fig. 6a). The whole brain was reviewed for scoring with researchers blinded to genotype.

Drug-inducible expression using geneswitch- (GS-) GAL4 drivers

0–2d adult flies were collected and aged for 24–48h prior to transfer to 0.04mg RU486-containing food (made by pipetting 100μl of 4mg/ml RU486 in 100% ethanol onto standard food and incubating at RT for 1–2d with slow rotation). Flies were transferred to fresh RU486-containing food every 2d.

Lifespan and negative geotaxis (climbing)

Lifespan and climbing assays were done as described with minor changes[23,61]. Assays were done at 24°C. For the climbing assay, flies were tapped to the bottom of an empty vial and immediately video recorded. Each vial was subjected to 3 consecutive trials with a 20–30min recovery between trials. For analysis, the number of flies in a vial that crossed a 4cm line after 15 or 20s post-tapping were counted. The mean value from the 3 trials per vial was used to account for technical variability.

Western immunoblots (WB)

Fly lysates and reagents: For assays involving Gmr-GAL4 (3d animals), triplicate samples of 5–10 heads per genotype were homogenized in 1X NuPAGE LDS sample buffer using disposable pellet/pestles tissue grinders (Kimble Chase #749520–0000) and motor (Kimble Chase #749540–0000). WBs were run using a standard protocol with Invitrogen's XCell SureLock blot system, 4–12% Bis-Tris NuPAGE gels, and a iBlot dry transfer system with nitrocellulose membrane. **iPSC lysates and reagents:** iPSC (described in Supplementary Fig. 10a) were cultured as previously described[67] and lysates were prepared using RIPA buffer. 20μg of total protein was run using a standard WB protocol with Invitrogen's XCell SureLock blot system, 4–12% Bis-Tris NuPAGE gels and wet transfer with PVDF membrane. **Mouse lysates and reagents:** Mice expressing either (G4C2)2 or (G4C2)149 were generated by intracerebroventricular administration of AAV2/9 vectors, previously described[51]. Cortical tissues were harvested at 3mo and 6mo. Protein lysates were prepped using a 1:5 weight/volume of ice-cold buffer (50mM Tris pH 7.4, 50mM NaCl, 1mM EDTA) with 2x protease and phosphatase inhibitors. 1% Triton X-100 and 2% SDS were added. Tissue homogenates were sonicated on ice and centrifuged at 4°C for 20min at 16,000xg. 30μg of total protein was run using a standard WB protocol with 10% Tris-Glycine gels and wet transfer with nitrocellulose membrane. **Fly antibodies**: anti-βgalactosidase (Promega #Z3781, 1:2,000), anti-TDP43 (ProteinTech #10782–2-AP, 1:2,000), anti-αTubulin (DSHB #AA4.3, 1:2,000)[23,35,54,61]. Antibodies targeting dCDC73 or dRtf1 were previously developed[68]. **Mammalian antibodies**: anti-Paf1 (Cell Signaling #12883, 1:1,000)[39], anti-Leo1 (ProteinTech #12281–1-AP, 1:2,000), anti-CDC73 (Cell Signaling #8126, 1:2,000)[69], anti-Rtf1 (Cell Signaling #14737, 1:2,000), anti-αTubulin-HRP (Cell Signaling #9099, 1:2,000), anti-GAPDH (IPSCs: Abcam # G8795, 1:5,000; Mice: Meridian Life Science # H86504M, 1:10,000). Antibodies validated by siRNA in fibroblast cells (data not shown): anti-Paf1, anti-Leo1, anti-Rtf1. **Secondary antibodies and imaging:** Mouse or Rabbit HRP-conjugated secondary antibodies (Jackson Immunoresearch Labs #115-035-146 and 111-035-144) were used at 1:5,000[23,35,54,61]. Blots were analyzed using Amersham ECL Prime Detection Reagent and imaged on an Amersham Imager 600.

Real-time quantitative PCR (qPCR)

All primers are defined in Supplementary Table 4. **Flies:** Total RNA was collected using TRIzol. For DaGS or Da-GAL4 assays, biological triplicate samples of 10 whole animals were processed per condition. For ElavGS assays, biological triplicate samples of 20–30 fly heads (16d) were processed per condition. qPCRs were run using standard protocols. cDNA was made from 200–400ng of total RNA using random primers, High Capacity cDNA Reverse Transcription Kit (ThermoFisher #4368814). qPCR reactions were setup using SYBR Green Fast reagents, 384-welled plate, and analyzed on an Applied Biosystem's ViiA 7 Real-Time PCR System. Mean fold change was determined using the Ct method. The housekeeping gene, RP49 (RNAPI-driven), was used. Primers targeting the G4C2 fly transgenes utilized unique restriction enzyme sequences in our construct found immediately 3' of the repeat. All primers were validated using a serial dilution curve. Supplementary Fig. 12 shows primer efficiencies between G4C2 and loading control used, RP49, to be similar. **Humans:** Total RNA was extracted from frozen postmortem tissue from the frontal cortex

using the RNAeasy Plus Mini Kit (QIAGEN), previously described[70]. RNA integrity (RIN) was verified on an Agilent 2100 bioanalyzer. qPCRs were run using standard protocols. cDNA was made from 500ng of total RNA (RIN 7.0, mean RIN for all samples =9.3) using random primers, High Capacity cDNA Reverse Transcription Kit (ThermoFisher). qPCR was conducted using SYBR GreenER qPCR SuperMix (Invitrogen) for all samples in triplicate. qPCRs were run in an ABI Prism 7900HT Fast Real-Time PCR System (Applied Biosystems). Fold change was determined using the   Ct method, comparing relative expression to healthy controls (mean=1). Housekeeping genes included RPLP0 and GAPDH. Primers targeting the *C9orf72* transcript were previously described[71]. **Yeast:** Centromeric galactose-inducible plasmids that express *C9orf72* hexanucleotide repeats – sense (G4C2)66 or antisense (G2C4)66 – were transformed into yeast, described[23]. Overnight cultures were grown from transformants in 2% raffinose-containing media, then diluted into 2% galactose-containing liquid media the next morning, and further grown at 30°C for 6 hours to allow for transgene expression. RNA was harvested from these cultures using a MasterPure Yeast RNA Extraction kit (Lucigen), including DNAseI digestions during the purification. Equal amounts of RNA were reverse transcribed using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) and analyzed by qPCR using SYBR green reagents.

External fly eye GR-GFP imaging

$y^1 sc^* v^1$;; Gmr-GAL4, UAS-LDS-(G4C2)44[GR-GFP] animals were crossed to RNAi lines or controls. Heads from 1–2d progeny were isolated and positioned for imaging on a glass slide using Vaseline. GFP imaging was immediately performed on a Leica DM6000B microscope using Z-stacks. Total GFP fluorescence in the compiled images was measured in ImageJ and normalized relative to control. Only expanded (G4C2)30+ transgenes produce GR-GFP; lines containing   22 repeats have no GFP signal by external eye fluorescence imaging or by western immunoblot.

Yeast strains and spotting assays

All yeast strains are derivatives of the haploid wildtype BY4741 strain, and all deletion strains were verified by PCR genotyping and qPCR. Centromeric galactose-inducible plasmids expressing codon-optimized *C9orf72* dipeptide repeat proteins – (PR)50 and (GR)100[23,72,73] – or a control protein (CCDB) were transformed into yeast using standard methods and selected for on SD-URA agar plates. For serial dilution growth analysis, transformants were grown overnight in 2% raffinose-containing media. Overnight cultures were all normalized to OD600 = 0.8, and then each strain was diluted serially 5-fold in a 96-well plate. Yeast were spotted onto either 2% glucose or 2% galactose containing agar plates with a multi-pin 'frogger' and allowed to grow for 48h at 30°C before photographed.

Fibroblast cells and chromatin immunoprecipitation (ChIP)

Primary fibroblast cells (described in Supplementary Fig. 10a) were cultured in a standard 37°C / 5% CO_2 incubator using DMEM (with L-GLUT, hi glucose and sodium pyruvate) supplemented with 15% heat-inactivated FBS, 1% amino-acids, and 1% pen/strep. For ChIP, cells were collected from a confluent T-75 flask using trypsin, washed in DPBS and fixed in

1.1% formaldehyde, 10min. ChIPs were then performed as described[74] with the following details. DNA was fragmented using a bucket sonicator for 20min. Dynabeads protein A (Invitrogen # 10001D) were used at 30μl per IP. The primary antibody against hLeo1 (ProteinTech #12281–1-AP) was used at 5μg for 100μg DNA, with overnight incubation. An overnight decrosslinking was done and DNA was isolated using a standard Phenol/Chloroform/Isoamyl alcohol (PCA) protocol into 50μl. 4μl was run per reaction by qPCR using Sybr green. Ct values were normalized to input and transformed relative to Rb IgG control (Cell Signaling #2729). Intergenic sequence was as described[75]. A positive control gene, p21, was used to validate Leo1 ChIPs[38]. All primers are defined in Supplementary Table 4.

Statistical analysis, sampling, and randomization

GraphPad Prism 7.00 or 8.00 software was used to develop all graphs and for all statistical analyses. P-values 0.05 were considered significant. See Supplementary Table 1 for details on statistics and Life Sciences Reporting Summary for additional information. No data were excluded from this study. **Flies, yeast, mice, and cell lines:** Data was assumed to be normal. A two-tailed unpaired student t-test, one-way ANOVA, or two-way ANOVA statistical analysis were performed when appropriate based on the experiment design. For ANOVA's Tukey's multiple comparisons tests were predominantly used. No statistical methods were used to predetermine sample sizes and numbers were similar to previous work[23,34,52,59,60]. Researchers were blinded to the genotype of all samples to maintain unbiased scoring. **Humans:** Data was found to not fit a normal distribution. Kruskal–Wallis, 1-way ANOVAs with Dunn's multiple comparisons test were performed to assess changes in expression of genes. Follow-up Spearman r correlations were performed. No statistical methods were used to predetermine sample sizes and numbers were similar to previous work[23,70,76]. Researchers were not blinded as data is inherently unbiased.

DATA AVAILABILITY STATEMENT

The authors are submitting all relevant data for publication. Any additional inquiries can be directed to the corresponding author, while any information relevant to this study will be openly shared including all raw data, unique reagents, or unique protocols.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

REFERENCES (main text)

1. DeJesus-Hernandez M et al. Expanded GGGGCC hexanucleotide repeat in non-coding region of C9ORF72 causes chromosome 9p-linked frontotemporal dementia and amyotrophic lateral sclerosis. Neuron 72, 245–256 (2011). [PubMed: 21944778]

2. Renton AE et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron 72, 257–268 (2011). [PubMed: 21944779]

3. van Blitterswijk M et al. Ataxin-2 as potential disease modifier in C9ORF72 expansion carriers. Neurobiol. Aging 35, 2421.e13–2421.e17 (2014).

4. van Blitterswijk M et al. TMEM106B protects C9ORF72 expansion carriers against frontotemporal dementia. Acta Neuropathol. (Berl.) 127, 397–406 (2014). [PubMed: 24385136]

5. Balendra R & Isaacs AM C9orf72 -mediated ALS and FTD: multiple pathways to disease. Nat. Rev. Neurol 14, 544 (2018). [PubMed: 30120348]

6. Yuva-Aydemir Y, Almeida S & Gao F-B Insights into C9ORF72-Related ALS/FTD from Drosophila and iPSC Models. Trends Neurosci (2018). doi:10.1016/j.tins.2018.04.002

7. Vatovec S, Kovanda A & Rogelj B Unconventional features of C9ORF72 expanded repeat in amyotrophic lateral sclerosis and frontotemporal lobar degeneration. Neurobiol. Aging 35, 2421.e1–2421.e12 (2014).

8. Vatsavayai SC, Nana AL, Yokoyama JS & Seeley WW C9orf72-FTD/ALS pathogenesis: evidence from human neuropathological studies. Acta Neuropathol. (Berl.) 137, 1–26 (2019). [PubMed: 30368547]

9. Rhodes D & Lipps HJ G-quadruplexes and their regulatory roles in biology. Nucleic Acids Res 43, 8627–8637 (2015). [PubMed: 26350216]

10. Simone R, Fratta P, Neidle S, Parkinson GN & Isaacs AM G-quadruplexes: Emerging roles in neurodegenerative diseases and the non-coding transcriptome. FEBS Lett 589, 1653–1668 (2015). [PubMed: 25979174]

11. Hall AC, Ostrowski LA, Pietrobon V & Mekhail K Repetitive DNA loci and their modulation by the non-canonical nucleic acid structures R-loops and G-quadruplexes. Nucleus 8, 162–181 (2017). [PubMed: 28406751]

12. Freudenreich CH R-loops: targets for nuclease cleavage and repeat instability. Curr. Genet 1–6 (2018). doi:10.1007/s00294-018-0806-z

13. Sauer M & Paeschke K G-quadruplex unwinding helicases and their function in vivo. Biochem. Soc. Trans 45, 1173–1182 (2017). [PubMed: 28939694]

14. Rondón AG, García-Rubio M, González-Barrera S & Aguilera A Molecular evidence for a positive role of Spt4 in transcription elongation. EMBO J 22, 612–620 (2003). [PubMed: 12554661]

15. Rondón AG, Gallardo M, García-Rubio M & Aguilera A Molecular evidence indicating that the yeast PAF complex is required for transcription elongation. EMBO Rep 5, 47–53 (2004). [PubMed: 14710186]

16. Zhou Q, Li T & Price DH RNA Polymerase II Elongation Control. Annu. Rev. Biochem 81, 119–143 (2012). [PubMed: 22404626]

17. Chen Y et al. DSIF, the Paf1 complex, and Tat-SF1 have nonredundant, cooperative roles in RNA polymerase II elongation. Genes Dev 23, 2765–2777 (2009). [PubMed: 19952111]

18. Hartzog GA & Fu J The Spt4-Spt5 complex: a multi-faceted regulator of transcription elongation. Biochim. Biophys. Acta 1829, 105 (2013). [PubMed: 22982195]

19. Jaehning JA The Paf1 complex: Platform or player in RNA polymerase II transcription? Biochim. Biophys. Acta BBA - Gene Regul. Mech 1799, 379–388 (2010).

20. Van Oss SB, Cucinotta CE & Arndt KM Emerging Insights into the Roles of the Paf1 Complex in Gene Regulation. Trends Biochem. Sci 42, 788–798 (2017). [PubMed: 28870425]

21. Liu C-R et al. Spt4 Is Selectively Required for Transcription of Extended Trinucleotide Repeats. Cell 148, 690–701 (2012). [PubMed: 22341442]

22. Cheng H-M et al. Effects on Murine Behavior and Lifespan of Selectively Decreasing Expression of Mutant Huntingtin Allele by Supt4h Knockdown. PLoS Genet 11, (2015).

23. Kramer NJ et al. Spt4 selectively regulates the expression of C9orf72 sense and antisense mutant transcripts. Science 353, 708–712 (2016). [PubMed: 27516603]

24. Porter SE, Washburn TM, Chang M & Jaehning JA The Yeast Paf1-RNA Polymerase II Complex Is Required for Full Expression of a Subset of Cell Cycle-Regulated Genes. Eukaryot. Cell 1, 830–842 (2002). [PubMed: 12455700]

25. Yang Y et al. PAF Complex Plays Novel Subunit-Specific Roles in Alternative Cleavage and Polyadenylation. PLoS Genet 12, (2016).

26. Fischl H, Howe FS, Furger A & Mellor J Paf1 Has Distinct Roles in Transcription Elongation and Differential Transcript Fate. Mol. Cell 65, 685–698.e8 (2017). [PubMed: 28190769]

27. Moniaux N et al. The Human RNA Polymerase II-Associated Factor 1 (hPaf1): A New Regulator of Cell-Cycle Progression. PLoS ONE 4, (2009).

28. Nguyen CT, Langenbacher A, Hsieh M & Chen J-N The Paf1 complex component Leo1 is essential for cardiac and neural crest development in zebrafish. Dev. Biol 341, 167–175 (2010). [PubMed: 20178782]

29. Wang P et al. Parafibromin, a Component of the Human PAF Complex, Regulates Growth Factors and Is Required for Embryonic Development and Survival in Adult Mice. Mol. Cell. Biol 28, 2930–2940 (2008). [PubMed: 18212049]

30. Bahrampour S & Thor S Ctr9, a Key Component of the Paf1 Complex, Affects Proliferation and Terminal Differentiation in the Developing Drosophila Nervous System. G3 GenesGenomesGenetics 6, 3229–3239 (2016).

31. Chaturvedi D, Inaba M, Scoggin S & Buszczak M Drosophila CG2469 Encodes a Homolog of Human CTR9 and Is Essential for Development. G3 GenesGenomesGenetics 6, 3849–3857 (2016).

32. Tan PPC, French L & Pavlidis P Neuron-Enriched Gene Expression Patterns are Regionally Anti-Correlated with Oligodendrocyte-Enriched Patterns in the Adult Mouse and Human Brain. Front. Neurosci 7, (2013).

33. Soutourina J Transcription regulation by the Mediator complex. Nat. Rev. Mol. Cell Biol (2017). doi:10.1038/nrm.2017.115

34. Mizielinska S et al. C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. Science 345, 1192–1194 (2014). [PubMed: 25103406]

35. Chung C-Y et al. Aberrant activation of non-coding RNA targets of transcriptional elongation complexes contributes to TDP-43 toxicity. Nat. Commun 9, (2018).

36. Chu X et al. Structural insights into Paf1 complex assembly and histone binding. Nucleic Acids Res 41, 10619–10629 (2013). [PubMed: 24038468]

37. Xu Y et al. Architecture of the RNA polymerase II-Paf1C-TFIIS transcription elongation complex. Nat. Commun 8, 15741 (2017).

38. Kim J, Guermah M & Roeder RG The Human PAF1 Complex Acts in Chromatin Transcription Elongation Both Independently and Cooperatively with SII/TFIIS. Cell 140, 491–503 (2010). [PubMed: 20178742]

39. Yu M et al. RNA polymerase II–associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. Science 350, 1383–1386 (2015). [PubMed: 26659056]

40. Mayer A et al. Uniform transitions of the general RNA polymerase II transcription complex. Nat. Struct. Mol. Biol 17, 1272 (2010). [PubMed: 20818391]

41. Mayekar MK, Gardner RG & Arndt KM The Recruitment of the Saccharomyces cerevisiae Paf1 Complex to Active Genes Requires a Domain of Rtf1 That Directly Interacts with the Spt4-Spt5 Complex. Mol. Cell. Biol 33, 3259–3273 (2013). [PubMed: 23775116]

42. Cao Q-F et al. Characterization of the Human Transcription Elongation Factor Rtf1: Evidence for Nonoverlapping Functions of Rtf1 and the Paf1 Complex. Mol. Cell. Biol 35, 3459–3470 (2015). [PubMed: 26217014]

43. Qiu H, Hu C, Gaur NA & Hinnebusch AG Pol II CTD kinases Bur1 and Kin28 promote Spt5 CTR-independent recruitment of Paf1 complex. EMBO J 31, 3494–3505 (2012). [PubMed: 22796944]

44. Dermody JL & Buratowski S Leo1 Subunit of the Yeast Paf1 Complex Binds RNA and Contributes to Complex Recruitment. J. Biol. Chem 285, 33671–33679 (2010). [PubMed: 20732871]

45. Amrich CG et al. Cdc73 Subunit of Paf1 Complex Contains C-terminal Ras-like Domain That Promotes Association of Paf1 Complex with Chromatin. J. Biol. Chem 287, 10863–10875 (2012). [PubMed: 22318720]

46. Xie Y et al. Paf1 and Ctr9 subcomplex formation is essential for Paf1 complex assembly and functional regulation. Nat. Commun 9, (2018).

47. Omer T et al. Neuroimaging patterns along the ALS-FTD spectrum: a multiparametric imaging study. Amyotroph. Lateral Scler. Front. Degener 18, 611–623 (2017).

48. Schönecker S et al. Atrophy in the Thalamus But Not Cerebellum Is Specific for C9orf72 FTD and ALS Patients – An Atlas-Based Volumetric MRI Study. Front. Aging Neurosci 10, (2018).

49. Gerlach JM et al. PAF1 complex component Leo1 helps recruit Drosophila Myc to promoters. Proc. Natl. Acad. Sci 114, E9224–E9232 (2017). [PubMed: 29078288]

50. Meehan TF et al. Disease Model Discovery from 3,328 Gene Knockouts by The International Mouse Phenotyping Consortium. Nat. Genet 49, 1231–1238 (2017). [PubMed: 28650483]

REFERENCES (methods)

51. Chew J et al. Aberrant deposition of stress granule-resident proteins linked to C9orf72-associated TDP-43 proteinopathy. Mol. Neurodegener 14, 9 (2019). [PubMed: 30767771]

52. McGurk L & Bonini NM Protein interacting with C kinase (PICK1) is a suppressor of spinocerebellar ataxia 3-associated neurodegeneration in Drosophila. Hum. Mol. Genet 21, 76 (2012). [PubMed: 21949352]

53. Burguete AS et al. GGGGCC microsatellite RNA is neuritically localized, induces branching defects, and perturbs transport granule function. eLife 4,

54. Mordes DA et al. Dipeptide repeat proteins activate a heat shock response found in C9ORF72-ALS/FTLD patients. Acta Neuropathol. Commun 6, (2018).

55. Yu Z et al. A fly model for the CCUG-repeat expansion of myotonic dystrophy type 2 reveals a novel interaction with MBNL1. Hum. Mol. Genet 24, 954–962 (2015). [PubMed: 25305073]

56. Ni J-Q et al. A genome-scale shRNA resource for transgenic RNAi in *Drosophila*. Nat. Methods 8, 405 (2011). [PubMed: 21460824]

57. Perkins LA et al. The Transgenic RNAi Project at Harvard Medical School: Resources and Validation. Genetics 201, 843–852 (2015). [PubMed: 26320097]

58. Dietzl G et al. A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. Nature 448, 151 (2007). [PubMed: 17625558]

59. Elden AC et al. Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. Nature 466, 1069–1075 (2010). [PubMed: 20740007]

60. Kim H-J et al. Therapeutic modulation of eIF2α-phosphorylation rescues TDP-43 toxicity in amyotrophic lateral sclerosis disease models. Nat. Genet 46, 152–160 (2014). [PubMed: 24336168]

61. Berson A et al. TDP-43 Promotes Neurodegeneration by Impairing Chromatin Remodeling. Curr. Biol 27, 3579–3590.e6 (2017). [PubMed: 29153328]

62. Eden E, Lipson D, Yogev S & Yakhini Z Discovering Motifs in Ranked Lists of DNA Sequences. PLoS Comput. Biol 3, (2007).

63. Eden E, Navon R, Steinfeld I, Lipson D & Yakhini Z GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 10, 48 (2009). [PubMed: 19192299]

64. Subramanian A et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci 102, 15545–15550 (2005). [PubMed: 16199517]

65. Supek F, Bošnjak M, Škunca N & Šmuc T REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. PLOS ONE 6, e21800 (2011). [PubMed: 21789182]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

66. Auluck PK, Chan HYE, Trojanowski JQ, Lee VM-Y & Bonini NM Chaperone Suppression of α-Synuclein Toxicity in a Drosophila Model for Parkinson's Disease. Science 295, 865–868 (2002). [PubMed: 11823645]

67. Kramer NJ et al. CRISPR–Cas9 screens in human cells and primary neurons identify modifiers of C9ORF72 dipeptide-repeat-protein toxicity. Nat. Genet 1 (2018). doi:10.1038/s41588-018-0070-7

68. Adelman K et al. Drosophila Paf1 Modulates Chromatin Structure at Actively Transcribed Genes. Mol. Cell. Biol 26, 250–260 (2006). [PubMed: 16354696]

69. Herr P et al. A genome-wide IR-induced RAD51 foci RNAi screen identifies CDC73 involved in chromatin remodeling for DNA repair. Cell Discov 1, 15034 (2015). [PubMed: 27462432]

70. Prudencio M et al. Distinct brain transcriptome profiles in C9orf72-associated and sporadic ALS. Nat. Neurosci 18, 1175–1182 (2015). [PubMed: 26192745]

71. Niblock M et al. Retention of hexanucleotide repeat-containing intron in C9orf72 mRNA: implications for the pathogenesis of ALS/FTD. Acta Neuropathol. Commun 4, (2016).

72. Jovičić A et al. Modifiers of C9orf72 DPR toxicity implicate nucleocytoplasmic transport impairments in c9FTD/ALS. Nat. Neurosci 18, 1226–1229 (2015). [PubMed: 26308983]

73. Chai N & Gitler AD Yeast screen for modifiers of C9orf72 poly(Glycine-Arginine) dipeptide repeat toxicity. FEMS Yeast Res doi:10.1093/femsyr/foy024

74. Lee TI, Johnstone SE & Young RA Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat. Protoc 1, 729–748 (2006). [PubMed: 17406303]

75. Kim N, Sun H-Y, Youn M-Y & Yoo J-Y IL-1β–specific recruitment of GCN5 histone acetyltransferase induces the release of PAF1 from chromatin for the de-repression of inflammatory response genes. Nucleic Acids Res 41, 4495–4506 (2013). [PubMed: 23502002]

76. Prudencio M et al. Repetitive element transcripts are elevated in the brain of C9orf72 ALS/FTLD patients. Hum. Mol. Genet 26, 3421–3431 (2017). [PubMed: 28637276]
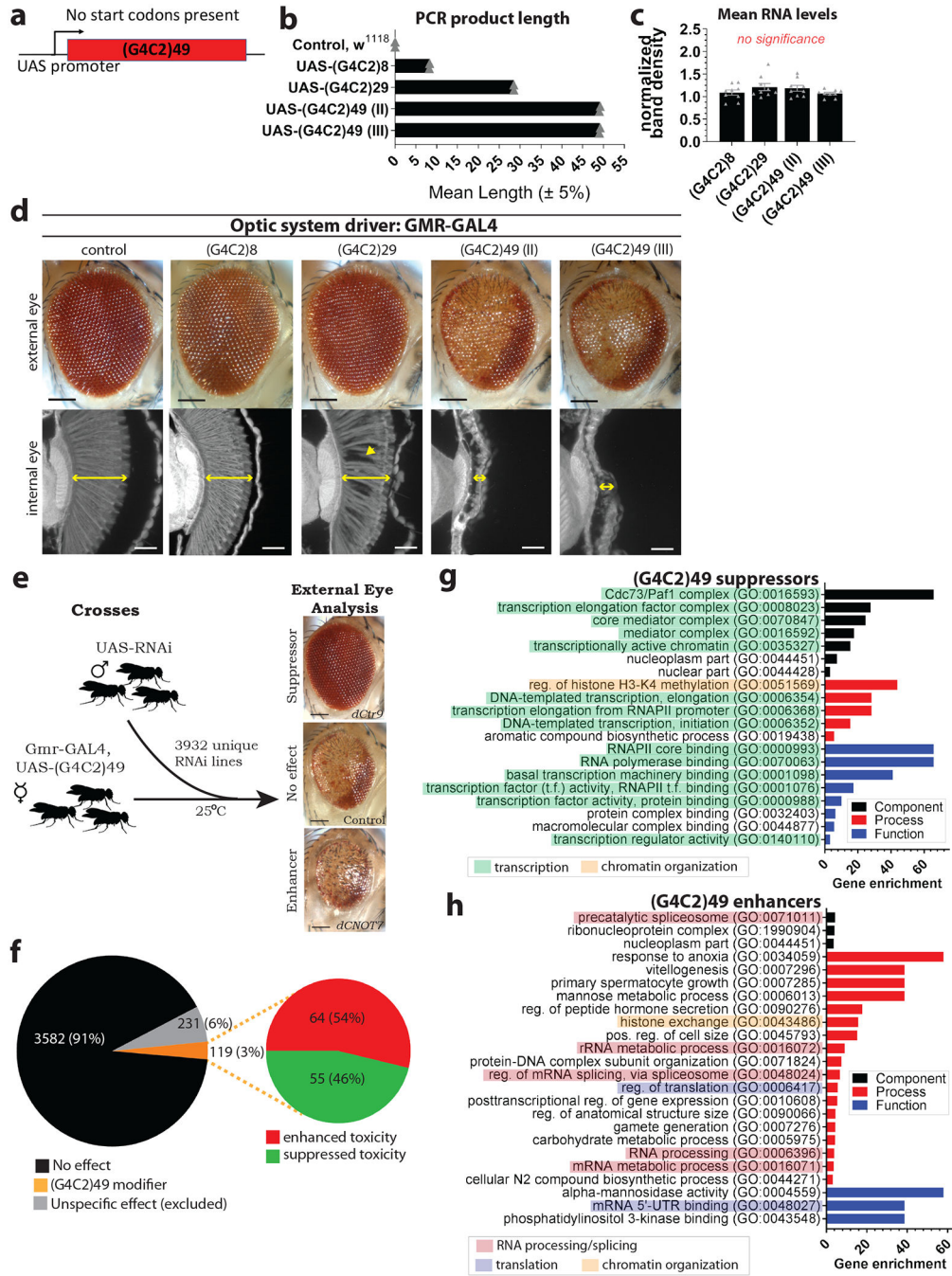
Figure 1: A genetic screen reveals PAF1C as a suppressor of (G4C2)49-toxicity in the fly eye.
(a) UAS-(G4C2)n transgenes were designed expressing a pure repeat. **(b)** PCR reactions were used to quantify of the number of repeats in individual UAS-G4C2 transgenic fly lines. Shown: individual data points with mean from 2 experiments (30 flies/experiment). **(c)** RNA expression of UAS-(G4C2)n transgenes using HS-GAL4 were compared by northern blots. Statistics: ANOVA with Tukey's correction, p-value: no significance >0.05. Shown: individual data points with mean±SEM; mean value of biological triplicates (n=30 flies) from 3 independent experiments. **(d)** Expression of UAS-(G4C2)n transgenes in the fly eye

compared to controls: (G4C2)8 had no effect, (G4C2)29 caused mild disruptions in 80% of animals, (G4C2)49 caused strong degeneration. Shown: data from one experiment; data reproduced in 3+ independent experiments. Arrows: internal tissue depth, lost tissue. **(e)** RNAi were co-expressed with (G4C2)49 (III) within the fly optic system. Effects of RNAi were recorded: "suppressors" reduced degeneration, "enhancers" increased degeneration. Shown: representative images. Hits were independently tested 3+ times to confirm reproducibility (>5 flies examined/cross). **(f)** 119 modifiers were identified. Control experiments excluded 231 RNAi lines with unspecific effects. **(g-h)** GO analyses revealed terms enriched in suppressors (55/3582 genes) or enhancers (64/3582 genes). Plotted: significant (p-value 10^{-3}) enrichment scores of >3.00. **(d-e)** Scale bars: external eye =100μm, internal eye =25μm. **(a-h)** Additional details for this and subsequent figures: Supplementary Fig. 1 (extended screen data), Supplementary Data (all screen results), Supplementary Table 1 (detailed sampling/reproducibility/statistics), methods.
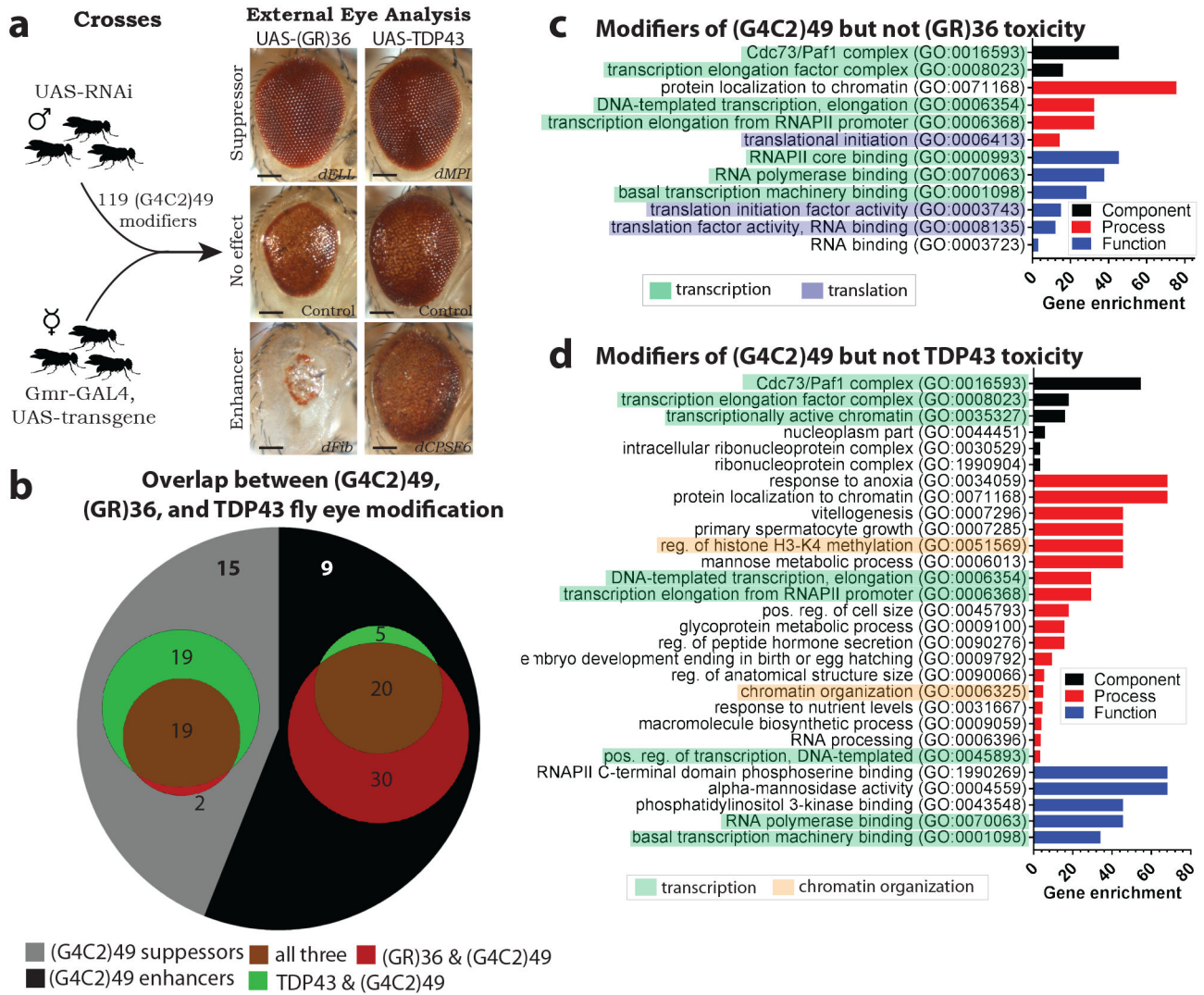
Figure 2: PAF1C is not a modifier of (GR)36 or TDP43 toxicity in *Drosophila*.
(a) The 119 (G4C2)49 modifiers were analyzed in (GR)36 and TDP43 models to determine if they could act on GR-dipeptide toxicity or had overlapping effects on TDP43-toxicity. All were independently tested 3+ times to confirm reproducibility of results (>5 flies examined/cross). Scale bars: 100μm. **(b)** Of the 119 modifiers of (G4C2)49 toxicity, 71 (59.7%) similarly modified (GR)36, arguing they may be acting on toxic DPR. 63 (52.9%) similarly affected TDP-43 toxicity, arguing overlap between these disease models. **(c)** GO analyses revealed terms enriched in the modifiers that *did not* similarly alter (GR)36 toxicity (48/3582 genes), revealing those acting selectively on the (G4C2)49 RNA model. **(d)** GO analyses revealed terms enriched in the modifiers that *did not* similarly alter TDP43 toxicity (56/3582 genes), revealing those specific to the expanded G4C2 repeat in C9+ FTD/ALS. **(c-d)** Plotted: GO-terms with significant (p-value   10^−3) enrichment scores of >3.00.
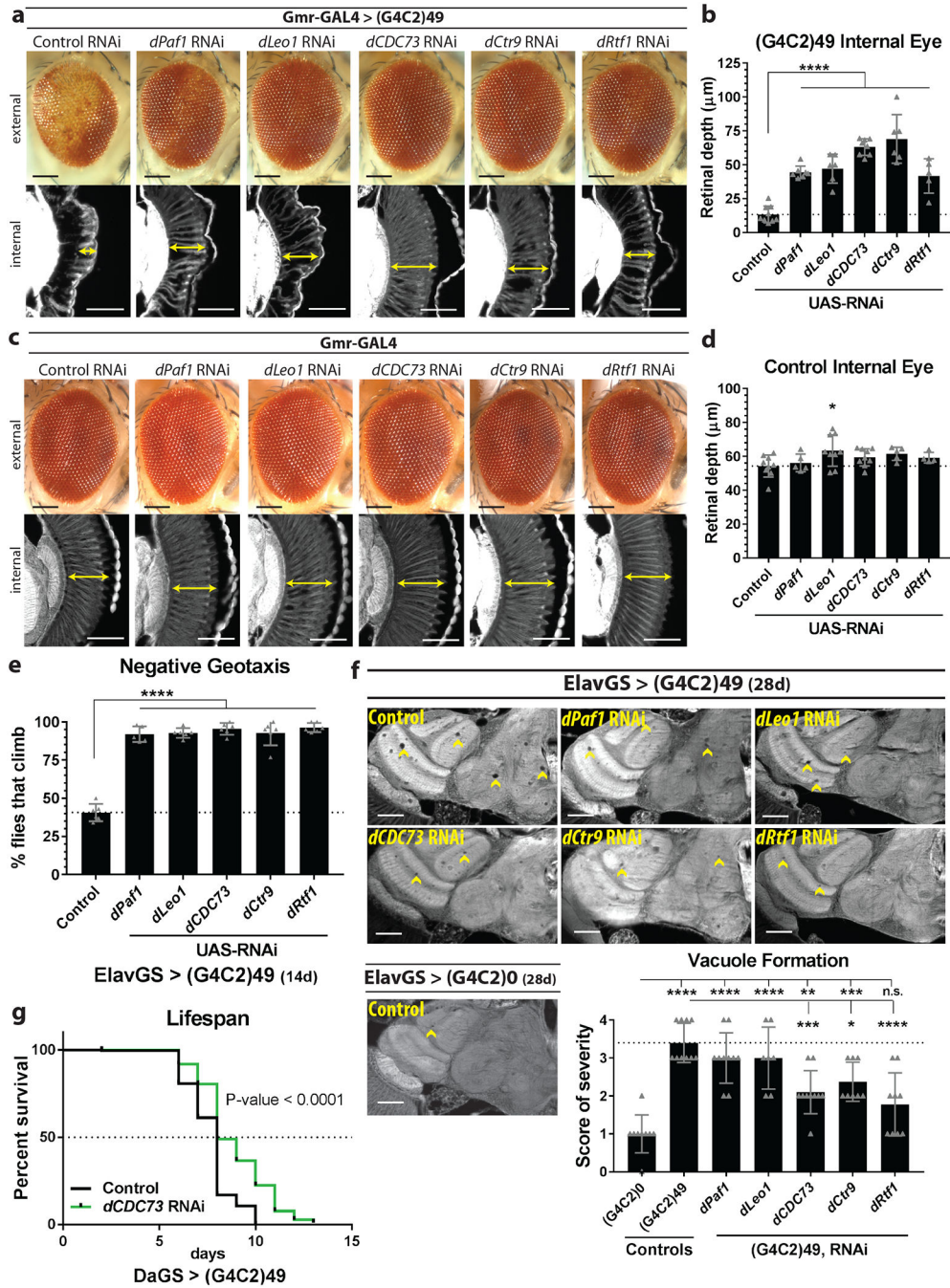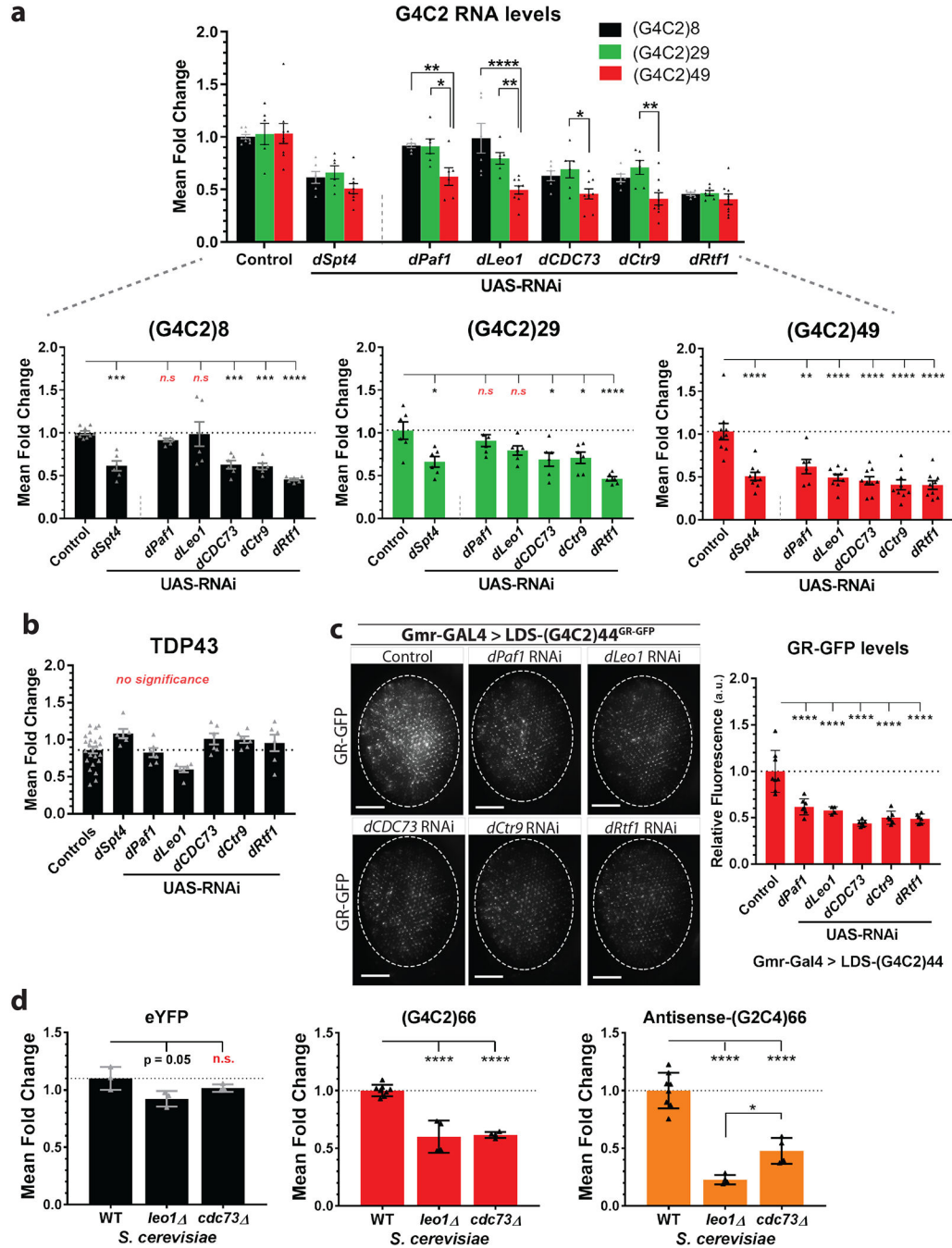
Figure 3: Reduced expression of components of PAF1C suppress (G4C2)49-induced toxicity in multiple contexts in the fly.

(a-b) dPAF1C RNAi mitigates toxicity associated with (G4C2)49 expression in the eye. N flies: control=9, *dPaf1*=7, *dLeo1*=6, *dCDC73*=8, *dCtr9*=6, *dRtf1*=5. **(c-d)** dPAF1C RNAi has no effect on control fly eyes. N flies: control=9, *dPaf1*=5, *dLeo1*=9, *dCDC73*=9, *dCtr9*=5, *dRtf1*=4. **(a-d)** Internal retina depth (arrows) quantified for individual animals. Scale bars: external eye =100μm, internal eye =45μm. **(e)** Climbing deficits caused by (G4C2)49 expression in the adult nervous system (ElavGS, 14d) are rescued by dPAF1C

RNAi. N flies: control=117, *dPaf1*=98, *dLeo1*=108, *dCDC73*=103, *dCtr9*=120, *dRtf1*=115. Individual data points are mean % of animals that could climb per tube; average of 19±2 animals per tube. **(f)** dPAF1C RNAi mitigated vacuole formation (arrowheads) in the brain with ElavGS driven expression of (G4C2)49. For quantification, a vacuole severity scoring system was developed where 0=no vacuoles and 4=medium/large, frequent (>5) vacuoles (see Sup. Fig. 6a). N flies: (G4C2)0=9, (G4C2)49: control=10, *dPaf1*=10, *dLeo1*=7, *dCDC73*=10, *dCtr9*=8, *dRtf1*=9. Scale bars: 50μm. **(g)** Knockdown of *dCDC73* in adult flies ubiquitously expressing (G4C2)49 results in lifespan extension. N flies: control=198, *dCDC73*=197. (b,d,f) each data point represents one animal. Statistics: (a-f) ANOVAs with Tukey's correction, (g) log-rank; p-values: ****<0.0001, ***<0.001, **<0.01, *<0.05, no significance (n.s.) >0.05. Shown on graphs: individual data points with mean±SD; data from one experiment; all experiments were repeated twice with similar results.

Figure 4: Downregulation of components of PAF1C selectively alter (G4C2)30+ transgene expression.

(a) (G4C2)n transgenes were co-expressed with dPAF1C RNAi lines in the adult brain using a drug inducible, neuronal driver (ElavGS, 16d). Transgene expression levels in heads measured by qPCR. *dPaf1* and *dLeo1* RNAi did not affect RNA levels of (G4C2)8 and (G4C2)29 but significantly reduced (G4C2)49 expression. *dSpt4*, *dCDC73*, *dCtr9*, *dRtf1* RNAi altered expression of all (G4C2)n transgenes. Shown: individual data points with mean±SEM; mean value of 3 biological replicates (n=25 flies/replicate) from 2–3

independent experiments. **(b)** dPAF1C RNAi did not alter TDP43 (ElavGS, 16d, heads) RNA levels by qPCR. Shown: individual data points with mean±SEM; mean value of 3 biological replicates (n=25 flies/replicate) from 2 independent experiments. **(c)** Downregulation of dPAF1C components reduces GR-GFP signal in LDS-(G4C2)44^{GR-GFP} animals. Quantification of total GFP fluorescence relative to control animals. N flies: control=8, *dPaf1*=7, *dLeo1*=4, *dCDC73*=6, *dCtr9*=6, *dRtf1*=6. Each data point represents one eye of one animal. Shown: data from one experiment; data reproduced in two independent experiments. Scale bars: 100μm. **(d)** Effects of deleting *scCDC73* (*cdc73*) or *scLeo1* (*leo1*) in *S. cerevisiae* on RNA levels from transgenes assessed by qPCR. Transgenes included eYFP (control), sense-(G4C2)66 (disease) and antisense-(G2C4)66 (disease). Shown: individual data points with mean±SD; mean value of biological duplicates from 2 independent experiments. Statistics: ANOVAs with Tukey's correction, p-values: ****<0.0001, ***<0.001, **<0.01, *<0.05, no significance (n.s.) >0.05.
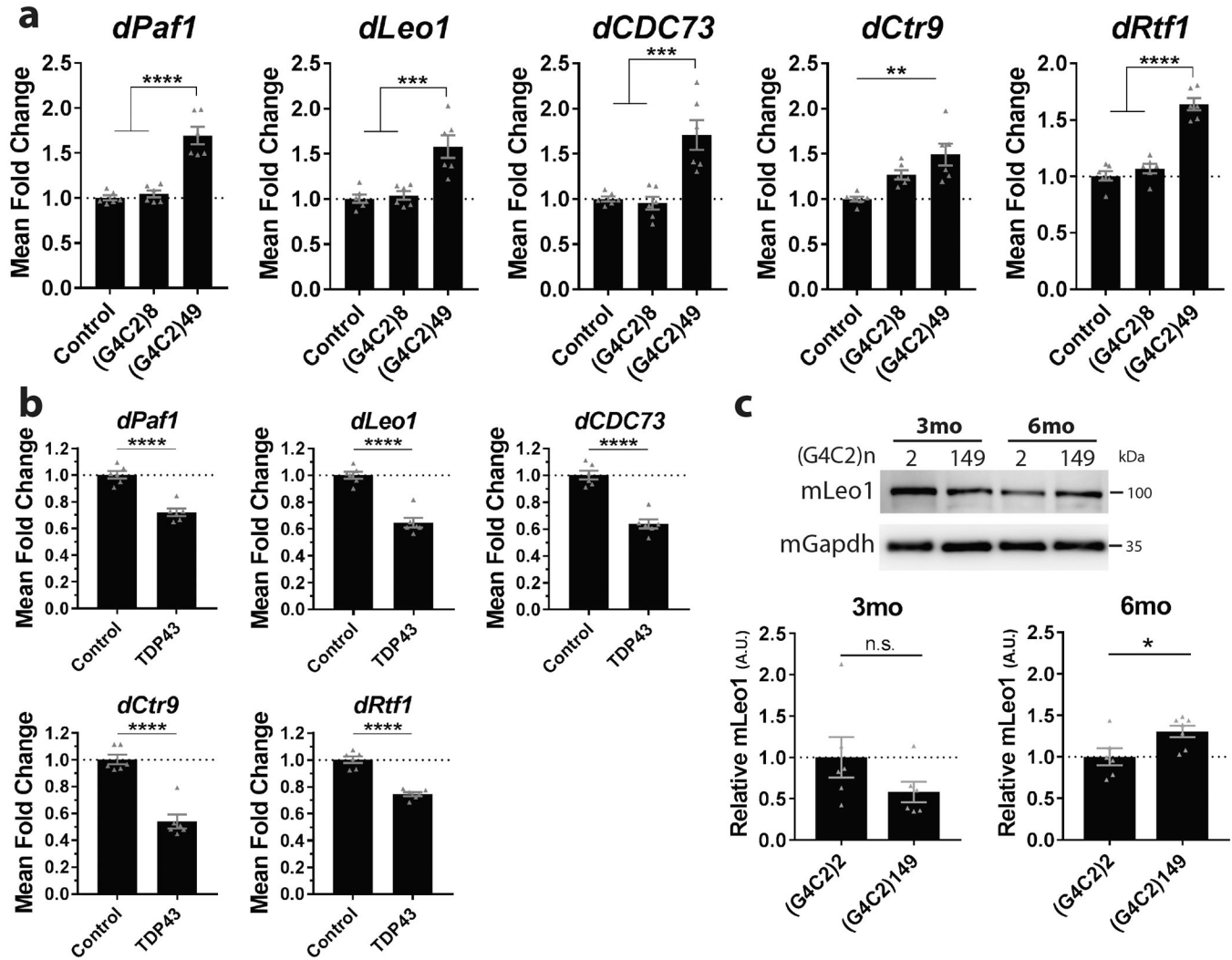
Figure 5: Endogenous PAF1C is upregulated in response to (G4C2)49 expression in the brain in flies and mice.

(a) Endogenous dPAF1C RNA expression is upregulated in (G4C2)49 expressing flies compared to control or (G4C2)8 expressing flies (qPCR). Transgenes expressed in the adult fly nervous system (ElavGS, 16d). Differences in expression are likely underestimated as RNA was extracted from whole head tissue, while transgenes were expressed selectively in neurons. **(b)** A non-G4C2 disease transgene, TDP43, was expressed with ElavGS (16d). Whole head analysis showed no upregulation of PAF1C components. **(a-b)** Shown: individual data points with mean±SEM; mean value of biological triplicates (n=25 flies/replicate) from 2 independent experiments. **(c)** Mouse endogenous mLeo1 protein levels measured in cortical tissue by western immunoblot using lysates from mice injected intracerebroventricularly with AAV2/9-(G4C2)2 or -(G4C2)149 at postnatal day 0. mLeo1 is upregulated by 6mo in response to expression of expanded (G4C2)149. Differences in expression are likely underestimated as protein was extracted from total cortical tissue while transgenes were expressed using AAV2/9 which predominantly transduces neurons. 3mo N animals: (G4C2)2=6, (G4C2)149=6. 6mo N animals: (G4C2)2=6, (G4C2)149=7. Shown:
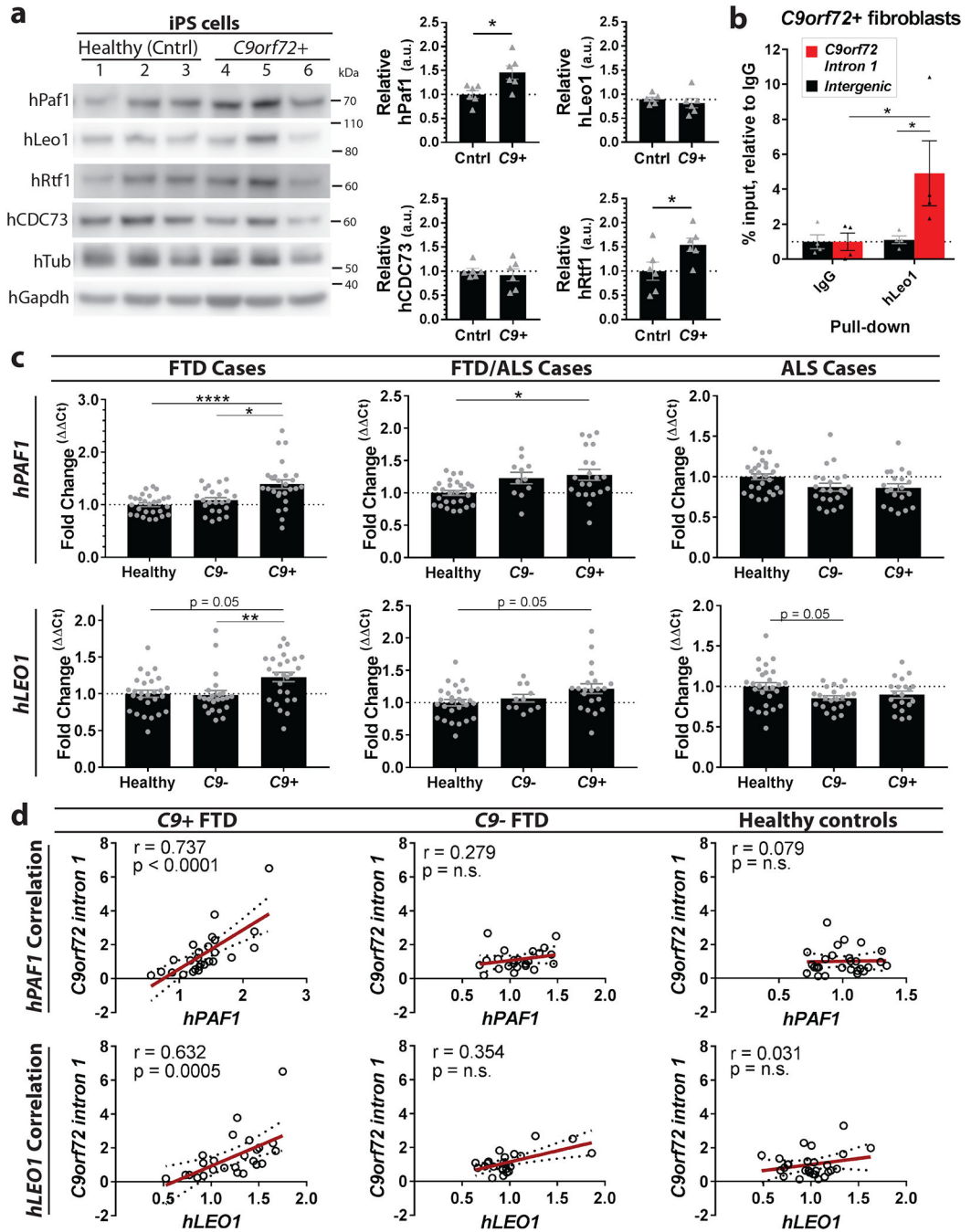
individual data points (each representing 1 animal) with mean±SEM. Data reproduced in two independent experiments. Statistics: (a) ANOVAs with Tukey's correction, (b-c) unpaired 2-tailed student t-test; p-values: ****<0.0001, ***<0.001, **<0.01, *<0.05, no significance (n.s.) >0.05. See Supplementary Figure 11 for uncropped western images for this and subsequent figures.

Figure 6: Upregulated *hPAF1* and *hLEO1* positively correlate to expression of repeat-containing *C9orf72* transcripts and hLeo1 binds *C9orf72*.
(a) Western immunoblots for hPAF1C components in iPS cells. hPAF1C band densities were normalized to the mean of loading controls: hTubulin, hGAPDH. Shown: individual data points with mean±SEM; mean value of 2 biological replicates from 2 independent experiments per cell line relative to the mean signal in controls. (b) Chromatin immunoprecipitation studies using a hLeo1 antibody on 4 independent *C9+*-derived fibroblast lines. Data: relative to IgG controls after normalizing by input. Shown: individual

data points (each representing 1 cell line) with mean±SEM; mean value of technical quadruplicates from 1 experiment. Data reproduced in 2 independent experiments per line. **(c)** qPCR analysis of *hPAF1* and *hLEO1* expression from healthy control (n=27), *C9*-(n=56), and *C9+* patients (n=67) frontal cortex tissue. Shown: individual data points (each representing 1 individual) with mean±SEM. **(d)** Spearman r coefficients defined correlations in expression from *hPAF1* or *hLEO1* expression with *C9orf72* transcripts in individuals. r values: 0=no correlation, 1.0=100% correlated. Shown: individual data points (each representing 1 individual) with linear regression±SE. Statistics: (a) unpaired 2-tailed student t-test, (b) ANOVA with Sidak's correction, (c) Kruskal-Wallis ANOVA with Dunn's correction, (d) Spearman R correlation; p-values: ****<0.0001, ***<0.001, **<0.01, *<0.05, no significance >0.05. *C9orf72 intron 1*: intronic region immediately 3' of the G4C2 repeat in the *C9orf72* gene. See Supplementary Figure 10 for cell line and patient characteristics.