



# RNA binding protein BOULE forms aggregates in mammalian testis

Yujuan Su<sup>1</sup>, Xinghui Guo<sup>1</sup>, Min Zang<sup>1</sup>, Zhengyao Xie<sup>1</sup>, Tingting Zhao<sup>1</sup>, Eugene Yujun Xu<sup>1,2,✉</sup>

<sup>1</sup>State Key Laboratory of Reproductive Medicine, Nanjing Medical University, Nanjing, Jiangsu 211166, China;

<sup>2</sup>Department of Neurology, and Center for Reproductive Sciences, Northwestern University, Chicago, IL 60611, USA.

## Abstract

Amyloids have traditionally been considered pathologic protein aggregates which contribute to neurodegeneration. New evidence however increasingly suggests that non-pathological amyloids are formed in animals during normal development. Amyloid-like aggregate formation was originally thought to be a conserved feature of animal gametogenesis. This hypothesis was based on findings which suggest that regulated amyloid formations govern yeast meiosis by way of meiosis-specific RNA binding proteins. Additional support came from studies which demonstrate that DAZL, a mammalian gametogenesis-specific RNA binding protein, also forms SDS-resistant aggregates *in vivo*. Here, we report evidence of aggregated BOULE formations, another DAZ family protein, during sperm development. Data suggest that in mouse testis, BOULE forms SDS-resistant amyloid-like aggregates. BOULE aggregate formation correlates with dynamic developmental expression during spermatogenesis but disappeared in *Boule* knockout testis. We also mapped essential small region *in vitro* BOULE aggregations, immediately downstream DAZ repeats, and found that aggregations positively correlated with temperature. We also performed enhanced UV cross-linking immunoprecipitation on BOULE aggregates from mouse testes and found that aggregates bind with a large number of spermatogenesis-related mRNAs. These findings provide insight into the amyloidogenic properties of gametogenesis-specific RNA binding proteins as a conserved feature in mammalian reproduction. Further investigation is warranted to understand the functional significance of BOULE amyloid-like formation during mouse spermatogenesis.

**Keywords:** amyloid, RNA-binding protein, BOULE, protein aggregation, SDD-AGE, enhanced UV cross-linking immunoprecipitation

## Introduction

Recent studies have shown that intracellular RNA-binding proteins (RBP) can play a role in neuronal development and through gametogenesis<sup>[1-2]</sup>, as well as during other processes. This is because RBPs

transform their physical state in order to generate functional aggregates<sup>[3-6]</sup>. RNA binding proteins in germ cells are also known to form membrane-less granules, also known as liquid droplets or condensates, in phase transitioning<sup>[7]</sup>. While physiological requirements and the function of human

✉ Corresponding author: Eugene Yujun Xu, State Key Laboratory of Reproductive Medicine, Nanjing Medical University, 101 Longmian Avenue, Nanjing, Jiangsu 211166, China. Tel: +86-25-86869505, E-mail: [e-xu@northwestern.edu](mailto:e-xu@northwestern.edu).

Received: 05 April 2022; Revised: 25 May 2022; Accepted: 27 May 2022; Published online: 28 June 2022

CLC number: R321.1, Document code: A

The authors reported no conflict of interests.

This is an open access article under the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited.

fertility factors are generally understood using mouse models<sup>[8]</sup>, little is actually known about how physical states and state transitions of those factors contribute to mammalian gametogenesis.

In yeast, the RNA-binding protein Rim4 mediates translational repression of numerous mRNAs by assembling them into amyloids. This shows that amyloid can be utilized as central regulators of gametogenesis and further, deletion of the low-complexity region prevents Rim4 amyloid formation and its ability to repress translation<sup>[1,9]</sup>. Similarly, Rim4, which is a member of the mammalian Deleted in Azoospermia (DAZ) protein family, is also a gametogenesis-specific RNA binding protein containing RNA recognition motif. The human DAZ protein family consists of Y-linked DAZ genes and autosomal homologs BOULE and DAZL, all of which function only in the testis or ovaries<sup>[10-12]</sup>. Human BOULE and DAZ proteins have a propensity to aggregate when expressed in yeast, and DAZ proteins are thought to contain a prion-like domain<sup>[13-14]</sup>. In addition, mouse DAZL can also form aggregates in mouse testis<sup>[1]</sup>; however, DAZ family protein aggregate-formation in mammalian spermatogenesis has not been explored. Given that BOULE is the oldest member of the DAZ family and is functionally conserved from fly to human<sup>[10,15]</sup>, we hypothesize that conserved BOULE might aggregate. Therefore, we chose BOULE to examine whether DAZ family proteins undergo amyloid-like aggregation during spermatogenesis. Such characterization may provide insights into mechanisms involved in DAZ protein regulation in mammalian fertility.

BOULE is conserved in sequences in the metazoan animal world and from sea anemones to humans. Mouse and fly *Boule* homologs also share a conserved spermatogenic requirement<sup>[10,15]</sup>. In a previous study, we found that *Boule* genes are lacking in mice which leads to male-specific infertility with sperm development arrested at the spermatid stage<sup>[10,16]</sup>. However, the specific mechanism involved in BOULE spermatogenesis regulation remains unclear. Based on the hypothesis that amyloid-like protein aggregation is a conserved feature of animal gametogenesis<sup>[1]</sup>, we proposed to determine whether BOULE could form amyloid-like aggregates during spermatogenesis in mice. We also looked to investigate BOULE functions through this evolutionarily conserved form as a functional amyloid-like aggregate.

Through the *in vivo* and the *in vitro* studies of mouse BOULE protein, we interrogated the aggregation propensity of mammalian BOULE protein, identified the regions essential for aggregation and established the functional significance of BOULE aggregation by performing eCLIP on aggregated

BOULE protein in the testis. Our data support that BOULE aggregation occurs during spermatogenesis as such their functional significance during mammalian spermatogenesis warrants further investigation.

## Materials and methods

### Animals

Wild type and *Boule* knockout mice were raised under standard conditions in the animal facilities of Nanjing Medical University, Nanjing, China. Protocols were approved by the Institutional Animal Care and Use Committees (IACUC) of Nanjing Medical University. All procedures were conducted according to institutional guidelines for the care and use of animals. Mice with a mixed 129/B6 background were back-crossed to C57/B6 over at least twenty generations with no changes in spermatogenic defects observed. All wild-type (WT) tissues, heterozygotes or homozygotes were from mice with mixed backgrounds from the same litter.

### Isolation of spermatogenic cells

Pachytene spermatocytes, round and elongated spermatids were isolated from adult mice using the STA-PUT method<sup>[17]</sup> with minor modifications. Testes were harvested and digested using collagenase type IV at 1 mg/mL. The dispersed seminiferous tubules were washed using Dulbecco's modified Eagle's medium (DMEM) (Gibco, USA) and centrifuged at 500 g. The pellet was then digested with 0.25% Trypsin (Gibco) containing DNase I (1 mg/mL; Sigma, USA) and filtered to prepare a single-cell suspension. Then, the single-cell suspension was loaded into a cell separation apparatus (ProScience, Canada), following a 2% to 4% bovine serum albumin (BEST, China) gradient. After 1.5 to 3 hours of sedimentation, cell fractions were harvested. The purity of different isolated cell types was determined according to morphological characteristics, cell diameter, and DAPI (Sigma) staining under a light microscope. The purity of spermatogonia, pachytene spermatocytes, and round spermatids is approximately 90%.

### Semi-denaturing detergent agarose gel electrophoresis assay

Semi-denaturing detergent agarose gel electrophoresis (SDD-AGE) was adapted from previous protocols described by Halfmann *et al*, with minor modifications<sup>[18-19]</sup>. Testis proteins and purified BOULE proteins were harvested and lysed in lysis buffer (50 mmol/L Tris-HCl, pH 7.5, 150 mmol/L NaCl, 0.5% Triton X-100, 10% glycerol with 1×

protease inhibitor cocktail [Roche, Switzerland]) for 30 minutes. Lysates were clarified twice through centrifugation at 5000 g for 5 minutes at 4 °C. 4× Sample loading buffer (2× TAE, 20% glycerol, 8% SDS, bromophenol blue) was added to lysates which were incubated for 10 minutes at room temperature, followed by loading with samples on newly-prepared 1.5% agarose gel with 0.1% SDS and electrophoresis in running buffer (1× TAE and 0.1% SDS) for 10 hours with a constant voltage of 29 V at 4 °C. Proteins were then transferred to the nitrocellulose filter membrane filter membrane (Merck Millipore, USA) for Western blotting with indicated antibodies by capillary action.

#### *Separation of Seprion PAD-beads*

The Seprion PAD-beads (SEP) isolation method is based on the proprietary ligand effect for retaining protein aggregates. A PAD-Bead kit (Microsens Biotechnologies, UK) was purchased and samples were treated in accordance with the manufacturer's protocol. Firstly, 25 mg of mouse testis tissue was homogenized in a total volume of 200 µL 10% sucrose. Then SDS was added to the homogenate to a final concentration of 0.1% (w/v) to enable solubilization. The captured protein was washed following the manufacturer's protocol and eluted at 95 °C for 5 minutes.

#### **Western blotting**

Tissues or cell samples were prepared by RIPA (Beyotime, China) and protease inhibitor cocktail. Aliquots of the lysates were electrophoresed in 10% polyacrylamide gel and transferred to polyvinylidene difluoride membrane (Bio-Rad, USA). The membrane was blocked with 5% skim milk/TBST (Tris-buffered saline with 0.1% Tween-20) and probed with various antibodies, *e.g.*, rabbit anti-BOULE (1:1000 dilution; Boule anti-serum 101)<sup>[16]</sup>, rabbit anti-PUM2 antibody (1:1000 dilution; Cat. No. ab10361, Abcam, USA); rabbit anti-PABP antibody (1:1000 dilution; Cat. No. ab21060, Abcam); rabbit anti- $\alpha$ -tubulin antibody (1:5000 dilution; Cat. No. SC-8035, Santa Cruz Biotechnology, USA); anti- $\beta$ -actin (1:1000 dilution; Cat. No. SC-1615, Santa Cruz Biotechnology); mouse anti-FLAG antibody (1:2000 dilution; Cat. No. abF1804, Sigma). Detection of HRP conjugated secondary antibody was performed with enhanced chemiluminescence detection reagents (ECL Kit; PerkinElmer, USA).

#### **Quantitative real-time PCR**

For quantitative real-time PCR (qPCR), the RNA

was converted to cDNA with random primers (Takara, Japan). qPCR was performed using a SYBR Green Master Mix Kit (Vazyme Biotech). Relative gene expression was analyzed based on the  $2^{-\Delta\Delta Ct}$  method with *Gapdh* as internal control. At least three independent experiments were analyzed. All primers were listed in the *Supplementary Table 1* (available online).

#### **Immunofluorescence**

The entire testes were fixed with Hartman's Fixative (Sigma), then paraffin-embedded and serially sectioned (5 µm) according to standard protocols<sup>[11,16]</sup>. Sections were used to perform immunofluorescence analysis. The primary antibodies used were as follows: rabbit anti-amyloid fibrils OC (1:1000 dilution; Cat. No. ab2286, Merck Millipore); rabbit anti-amyloid oligomer/rabbit anti-A11 (1:1000 dilution; Cat. No. ab9234, Merck Millipore); rabbit anti-BOULE (1:200 dilution; Boule anti-serum 101)<sup>[16]</sup>.

#### **Expression and purification of recombinant BOULE proteins**

The bacterial expression vector pET28a-His6-mBOULE-FL (full-length) was transformed into BL-21. Protein expression was induced with 0.5 mmol/L IPTG (Beyotime) at 16 °C for 20 hours. After sonication in lysis buffer (50 mmol/L Tris-HCl [pH 8.0], 150 mmol/L NaCl, 10% glycerol, 0.1% TritonX-100, 1 mmol/L PMSF, and 10 mmol/L imidazole), cell lysates were centrifuged at 20 000 g for 30 minutes. His6-mBOULE-FL in the supernatant were purified using Ni-NTA spin column (Bio-Rad), washed with wash buffer (50 mmol/L Tris [pH 8.0], 150 mmol/L NaCl, 10% glycerol, 0.1% TritonX-100, 1 mmol/L PMSF, and 20 mmol/L imidazole) and eluted with elution buffer (50 mmol/L Tris [pH 8.0], 150 mmol/L NaCl, 10% glycerol, 0.1% TritonX-100, 1 mmol/L PMSF, and 200 mmol/L imidazole), then verified using Coomassie Brilliant Blue staining (*Supplementary Fig. 1*, available online). Fractions containing His6-mBOULE-FL were pooled and loaded onto a Amicon Ultra-4 filtration column (Merck Millipore) equilibrated with lysis buffer (50 mmol/L Tris-HCl [pH 7.5], 150 mmol/L NaCl, 0.5% TritonX-100, 1× cocktail).

To verify the effect of different temperatures and concentrations on purified BOULE protein, concentrated BOULE solution was serially diluted into 9 mg/mL (280 mmol/L), 0.9 mg/mL (28 mmol/L), 0.09 mg/mL (2.8 mmol/L), the purified protein of each concentration was evenly divided into three parts, and placed at 4 °C, 25 °C and 37 °C for 30 minutes, and then followed by the SDD-AGE assay.

### Plasmid construction and transfection

For BOULE expression using the pCMV6 vector, segments of the mouse BOULE were subcloned into the pCMV6 vector using *SgfI* and *MluI* restriction enzymes (New England BioLabs Inc., USA). pCMV6-mouse BOULE deletion plasmids (Del-1 to Del-5) were generated by PCR amplification of the coding sequences (CDS) followed by recombining into pCMV6-Entry destination plasmid (Origene Inc., USA) using the ClonExpress MultiS One Step Cloning Kit (Vazyme Biotech). HEK293T cells in a 6-well plate were transfected with 5 µg of pCMV6 construct carrying either segment of BOULE using ExFect Transfection Reagent (Vazyme Biotech). After 48 hours, cells were harvested for SDD-AGE analysis according to the manufacturer's instructions.

For BOULE expression using pCDNA3.1, segments of the mouse BOULE were subcloned into the pCDNA3.1 vector by using *HindIII* and *BamHI* restriction enzymes (New England BioLabs Inc.). Plasmid pCDNA3.1-EGFP-mBOULE-FL and pCDNA3.1-EGFP-mBOULE-Delc were generated by PCR amplification of the CDS followed by recombining into pCDNA3.1-Entry destination plasmid (Invitrogen, USA) using the ClonExpress MultiS One Step Cloning Kit. HeLa cells in a 3.5 cm glass dish plate were transfected with 4 µg of pCDNA3.1 construct carrying either segment of BOULE using ExFect Transfection Reagent. Droplets were observed 24 hours later on a glass-bottom cell culture dish for fluorescence imaging.

### Fluorescence recovery after photobleaching analysis

EGFP-mBOULE-FL proteins were used to label liquid droplets *in vitro*. We performed fluorescence recovery after photobleaching (FRAP) assaying in HELA cells and FRAP experiments were performed on a confocal microscope (LSM 800, Zeiss) at room temperature. Defined regions were photobleached at 488 nm and the fluorescence intensities in these regions were collected every 1 second for *in vitro* droplets and normalized to the initial intensity before bleaching. Image intensity was measured in the region of interest.

### *BOULE aggregates enhanced UV cross-linking immunoprecipitation*

BOULE aggregates enhanced UV cross-linking immunoprecipitation (eCLIP) was performed as previously described<sup>[20]</sup>. This eCLIP protocol started with UV crosslinking of seminiferous tubules from mouse testes on postnatal day (P) 25, followed by partial RNase digestion and immunoprecipitation using an anti-BOULE polyclonal serum. RNA-protein

extracts from input, and from BOULE-IP of both wild-type testes and *Boule* knocked-out testes were eluted with an acidic elution buffer (pH 2.0 at 25 °C and later neutralized by pH 9.8 Tris buffer), based on non-denaturing elution (Dynabeads Protein G; Cat. No. 10004D, Thermo Fisher Scientific, USA), and was run by SDD-AGE and transferred to nitrocellulose filter membrane, specific regions (250 kDa or bigger bands) were excised from BOULE-IP lane. Only regions corresponding to aggregated BOULE were excised at molecular weight of 250 kDa and higher while leaving lower molecular weight corresponding to the BOULE monomers untouched. The region from 75 kDa or higher from both the input lane and *Boule* KO lane was then excised for library construction and sequence analysis as background control. This acidic elution step of the beads and SDD-PAGE are the main difference between aggregates eCLIP and regular eCLIP. The excised protein-bound RNAs were treated with proteinase K to extract aggregate-bound RNAs. After reversed transcription, the adapter is ligated to its 5' end. Then, the index primer was ligated to cDNA followed by PCR amplification. We sequenced a number of subclones to confirm the quality of libraries before high-throughput library sequencing.

### Bioinformatics analysis

#### *Sample demultiplexing and trimming*

Samples of IP and Input were demultiplexed by inline barcode<sup>[20]</sup>. Then, the 3' adapters and 5' adapters of raw reads were trimmed using in-house parameters which run two rounds according to library construction and structures of sequenced reads.

#### *Removing repetitive elements and mapping*

Trimmed reads were mapped to RepBase (Mouse RepeatMasker, version mm10) using STAR software to control for spurious artifacts from rRNA and other repetitive reads. Then, unmapped reads from the last step were mapped to the *Mus musculus* genome (UCSC, mm10).

#### *Peak calling*

After removing the repetitive elements and mapping with the mouse genome, the uniquely mapping reads were followed the algorithm field to remove eCLIP PCR duplicates. Then, we used tag2peak.pl in the CLIP Tool Kit, which calculates the number of overlapping CLIP tags in each genomic position to identify peaks for the next normalization.

#### *Normalization with biological input*

BOULE eCLIP peaks of two replicates were



normalized by 2 corresponding Inputs *via* a script<sup>[20]</sup> to remove background noise signals. After that, we set different fold change lines to select significant peaks through normalization controls for further analysis.

### Statistical analysis

Statistical analyses were performed using GraphPad Prism 8.3.0 (GraphPad, USA). All quantitative experiments were repeated three times. The data were expressed as mean±SD. Two-group comparisons were performed using two-tailed unpaired *t*-test. The threshold for statistical significance was set at  $P<0.05$ .

## Results

### BOULE formed SDS-resistant aggregates in the testis

Since DAZL can form aggregates<sup>[1]</sup>, we asked whether the DAZL paralog, BOULE protein can also form aggregates. Many neurodegenerative-disease proteins and yeast proteins form fibrillar amyloids which can be seen using SDD-AGE<sup>[2,9,18,21–22]</sup>. Using this same method, we found a smear of SDS-resistant high-molecular-weight BOULE aggregates which were higher than 250 kDa in mice testis, and several times larger than the predicted 37 kDa, 45 kDa, and 55 kDa of BOULE monomer (**Fig. 1A**).

Given that murine BOULE was expressed in spermatogenic cells from pachytene to round spermatids<sup>[12]</sup>, we asked at what stage of spermatogenesis BOULE aggregates could be detected. Taking advantage of the synchronous spermatogenic development during the first wave of spermatogenesis in neonatal mice, we examined the expression of BOULE aggregates in the testes from P7, P14, P21, P35, and in adult mice, BOULE aggregates were detected first in P14 testis at a very low level. However, this increased significantly in P21 testes and persisted through and into adult testis (**Fig. 1A**). In P14 testis, spermatogenesis reached the early pachytene stage, suggesting that BOULE aggregate expression starts in early pachytene spermatocytes. A high molecular weight smear, detected using the BOULE antibody, was absent in the testis of *Boule* knockout mice<sup>[16]</sup> and wild-type brain tissue where there is no BOULE expression (**Fig. 1A**). This appears to confirm that these aggregates are specific to BOULE proteins.

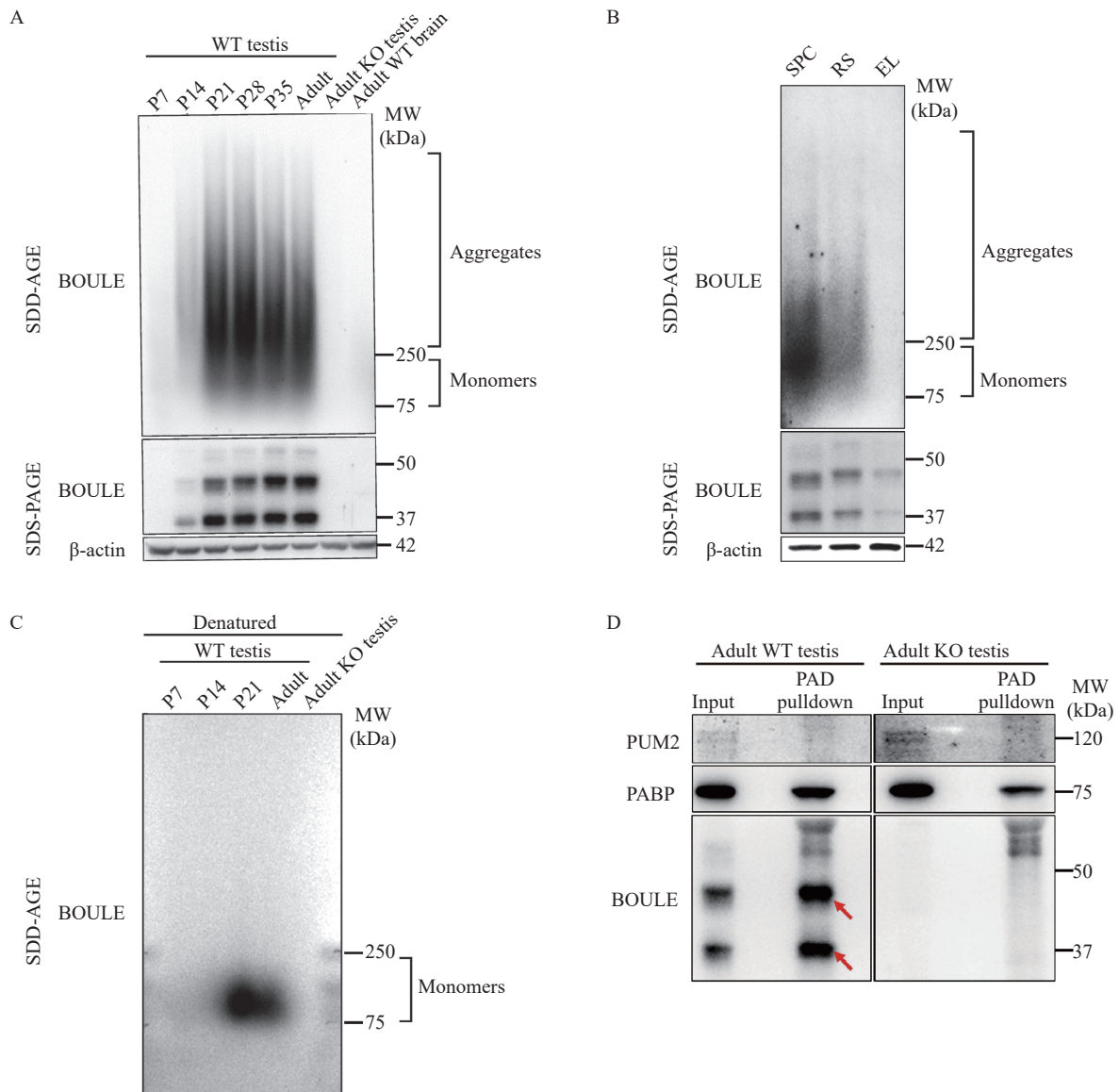
We also performed standard Western blotting analysis of proteins extracted from the same testes using SDS-PAGE to detect BOULE monomers. BOULE aggregates detected using SDD-AGE and BOULE monomers detected by SDS-PAGE exhibited

similar expression dynamics, low at P14 and high from P21 to adult (**Fig. 1A**). Concordant expression between BOULE monomers and aggregates was also observed in purified spermatogenic cells under the STA-PUT method. This supports the notion that BOULE aggregates are expressed specifically in spermatocytes and in round spermatids (**Fig. 1B**). A major difference between SDD-AGE and SDS-PAGE was the sample preparation, samples for SDD-AGE were not heat-denatured and there was no reducing agent ( $\beta$ -mercaptoethanol) in SDD-AGE lysis buffer. We prepared mouse testis protein extract samples by adding  $\beta$ -mercaptoethanol and denaturation them by heating, which resulted in complete denaturation of protein aggregates. We found that only BOULE protein signals at sizes smaller than 250 kDa but bigger than 75 kDa were detected by SDD-AGE gel, representing BOULE monomers on SDD-AGE gel (**Fig. 1C**). This experiment further supports the hypothesis that high molecular weight BOULE aggregates under SDD-AGE are indeed aggregates of many BOULE monomers.

### BOULE aggregates formation was confirmed by PAD assay and co-localized with fibril proteins during sperm development

To further verify the existence of BOULE aggregates, we used a commercially available protein aggregate binder *i.e.*, SEP on the testis BOULE aggregates<sup>[23]</sup>. We found BOULE proteins are enriched by SEP from wild type testes but not in the testis of *Boule* KO mice. This further supports the idea that BOULE can form aggregates (**Fig. 1D**). PAD is widely used in commercial settings and amyloid disease diagnoses such as bovine spongiform encephalopathy (aka mad cow disease) by pulling down cross beta-sheet amyloid. The successful pull-down of BOULE protein aggregates by PAD supported BOULE could therefore form amyloid-like aggregates.

We next investigated whether BOULE protein expression and aggregates expression occur during spermatogenesis by immunofluorescence staining of BOULE and with fibrillar and oligomeric amyloids extracted from mouse testis. Two conformation-specific antibodies A11 and OC, which recognize prefibrillar oligomeric and fibrillar forms of amyloid, respectively<sup>[24]</sup>, were used to visualize the expression of amyloids during spermatogenesis. Interestingly, BOULE signals colocalized with both A11 and OC signals in the testis (**Fig. 2A** and **B**). Specifically, BOULE and the aggregates are highly expressed in spermatocytes and round spermatids but appear absent in spermatogonia or elongating spermatids (**Fig. 2**).



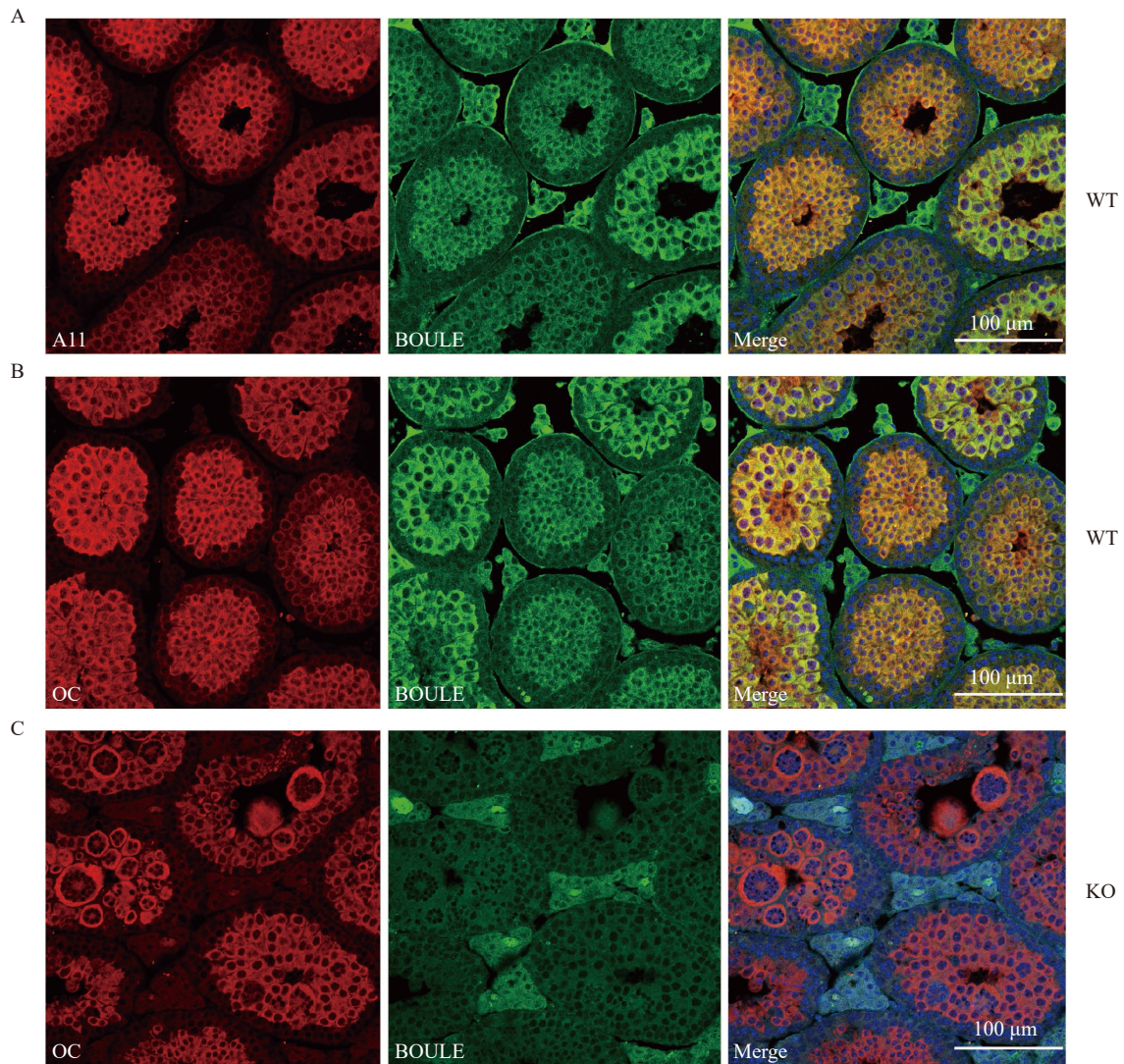
**Fig. 1** Mouse BOULE formed SDS-resistant aggregates in the testis. A: Testes at postnatal day (P) 7, P14, P21, P28, P35 of the first wave of spermatogenic from wild type (WT) mice as well as adult wild type testes and *Boule* knockout (KO) testes were homogenized and analyzed by SDD-AGE and SDS-PAGE assays with anti-BOULE antibody. Mouse brain protein extract was added as a negative control. B: Purified spermatogenic cells *via* the STA-PUT method were homogenized and analyzed by SDD-AGE and SDS-PAGE assays. SPC: spermatocyte; RS: round spermatids; EL: elongating spermatids. C: Testes at P7, P14, P21 representing different time points during the first wave of spermatogenesis from postnatal wild type mice as well as adult wild type testes and *Boule* knockout testes were homogenized in SDS-PAGE buffer (to be denatured) and then analyzed by SDD-AGE assays with anti-BOULE antibody. D: Testes from adult wild type mice as well as adult *Boule* knockout mice were assayed for BOULE aggregates with a commercially available protein aggregate binder, Seprion PAD-beads, *via* BOULE antibody pull down. PABP protein is known to form aggregates and acted as positive control while PUM2 protein was the negative control.

However, we also detected OC signals in *Boule* KO mice which suggests that BOULE is not the only protein in the testis that forms amyloid aggregates (**Fig. 2C**). Collectively, our data revealed that BOULE can form SDS-resistant high molecular amyloid-like aggregates in mouse testis.

### Key regions essential for the formation of BOULE aggregates

It is understood that proteins commonly contain low

complexity domains (LCDs) or intrinsically disordered regions (IDRs) that can drive full-length proteins into liquid-liquid phase separation (LLPS) through weak multivalent interactions<sup>[7,25–26]</sup>. It is also held that this can promote the formation of aggregates. Phase transition or condensate could be early-stage of BOULE aggregate formation. Therefore, we investigated this scenario and found several IDRs in mouse BOULE protein *via* MobiDB (**Fig. 3A**). To determine which region of BOULE proteins is



**Fig. 2** Mouse BOULE protein co-localized with amyloid structures during spermatogenesis. A: Immunofluorescence (IF) staining to detect the specific localization of BOULE was performed using anti-BOULE antibody and conformation-specific anti-A11 antibody in wild type (WT) mouse testis. BOULE (green) were colocalized in the cytoplasm with A11-positive signals (red). B: IF staining of WT mouse testis sections using anti-BOULE antibody and conformation-specific anti-OC antibody (red) showed BOULE proteins (green) colocalized with OC-positive signals (red) in the cytoplasm of spermatocytes and round spermatids. C: IF staining showed *Boule* knockout (KO) testis retained anti-OC signals (red). The giant round structures inside the seminiferous tubules are symplastic spermatids associated with spermatid arrest phenotype in *Boule* knockout testes. Scale bars: 100  $\mu\text{m}$ .

involved in aggregate formation, we constructed five different FLAG-tagged BOULE deletion constructs (Del-1, Del-2, Del-3, Del-4, and Del-5). Using SDD-AGE, we found those lacking Del-5 segment 5 (*i.e.*, amino acids 190–261) in BOULE protein failed to form aggregates (**Fig. 3B** and **C**). Therefore, amino acids 190 to 261 contain sequences which are vital when forming SDS-resistant aggregate structures (**Fig. 3B** and **C**).

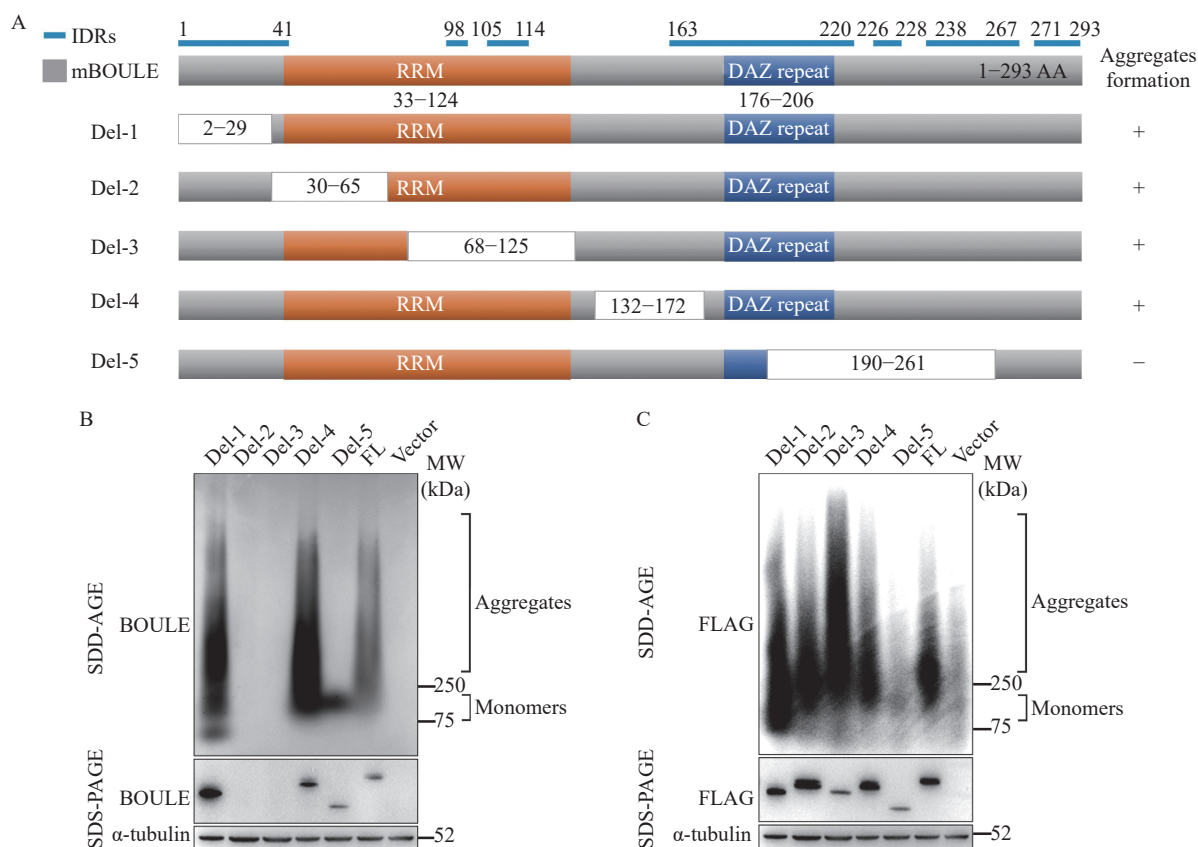
To further specify the region, we constructed six smaller deletions of the segment 5 region (Del-a, -b, -c, -d, -e, -f). We found that removal of the Del-c region (amino acids 210–240) of BOULE protein creates a deficiency in aggregation-forming abilities

(**Fig. 4B**). This enabled us to hone in on the BOULE region to understand the formation of BOULE aggregates to amino acids 210 to 240, consisting of VPQSPASSAPFLYLQPSEVIYQPVEIAQDGG. This small segment of peptides also contains sequences involved in an intrinsically disordered region (**Fig. 4A**) and is highly conserved in BOULE homologs of many mammalian species (data not shown). The generation of knock-out mutations which disrupt this region may provide insight into the function of BOULE aggregation in spermatogenesis.

#### *BOULE* aggregation was susceptible to heat shock

Since we found that BOULE aggregates in the testis





**Fig. 3** BOULE protein deletion constructs were used to map the region of BOULE essential for protein aggregation. A: Diagram of BOULE deletion fragments used. The blue line is the intrinsically disordered regions (IDRs) identified *via* MobiDB (<https://mobidb.bio.unipd.it>). B: pCMV6-mouse *Boule* deletion plasmids (Del-1 to Del-5), as well as pCMV6 mouse BOULE FL plasmid, were transfected into HEK293T cells, respectively, and plasmid pCMV6-vector only was added as a negative control. After 48 hours, cells were harvested for SDD-AGE and SDS-PAGE assays to detect BOULE aggregates and monomers with anti-BOULE antibody and anti- $\alpha$  Tubulin antibody. C: BOULE aggregates and monomers were verified by SDD-AGE and SDS-PAGE with anti-FLAG and anti- $\alpha$ -tubulin antibodies. RRM: RNA Recognition Motif; DAZ: Deleted in Azoospermia.

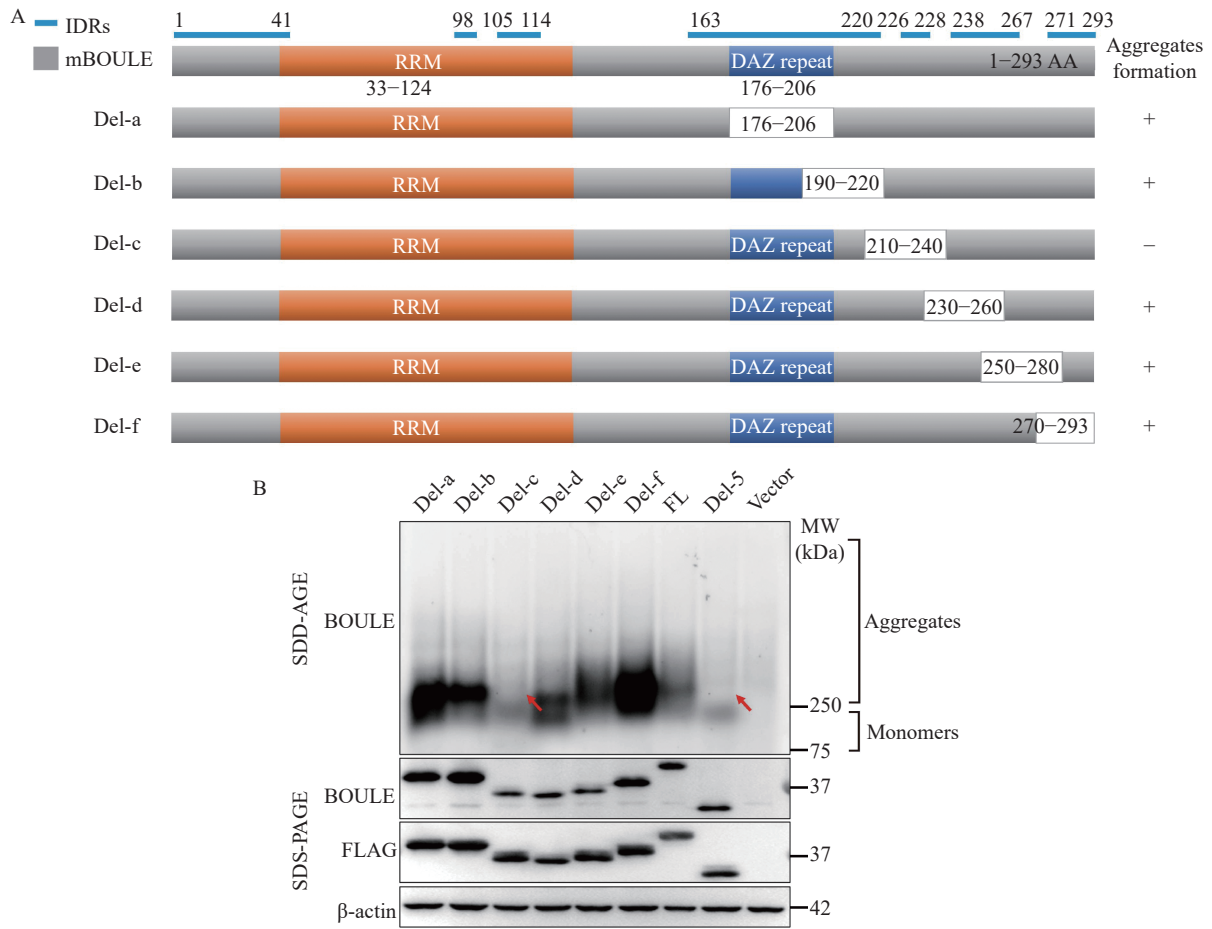
and that 210–240 AA of BOULE protein is crucial for the aggregation state of BOULE, we investigated whether BOULE can aggregate *in vitro*. To answer this question, we expressed 6 $\times$  His-tagged mouse BOULE full-length proteins in bacteria and purified the proteins (**Supplementary Fig. 1**, available online). Purified mouse BOULE incubated at different temperatures and different concentrations were analyzed by SDD-AGE assay to determine whether purified BOULE protein could form aggregates. We found that at 0.9 mg/mL (28 mmol/L) and higher, BOULE does clearly form protein aggregates. Higher temperatures promoted purified BOULE protein to form aggregates with high molecular weights (**Fig. 5A**), though at a concentration of 0.09 mg/mL (2.8 mmol/L) BOULE aggregation is still detectable. To exclude the potential impact of His-tag on BOULE protein aggregation property, we constructed N-terminal, C-terminal full-length BOULE protein together with *Boule* deletion construct (Del-c) and our data (**Supplementary Fig. 2**, available online) further

supported that BOULE full-length protein formed aggregates.

We also examined the effect of heat stress on BOULE aggregate formation and found that heat shock treatment of mouse testes at 42  $^{\circ}$ C increased BOULE aggregation in mouse testes compared to non-heated, shocked mice (**Fig. 5B**). This is consistent with the previous report of DAZ family proteins in stress granules<sup>[27]</sup>.

Next, we constructed BOULE FL, BOULE protein 210–240 amino acids deletion construct (BOULE Del-c), and enhanced green fluorescent protein (EGFP) fusion constructs and transfected them into HeLa cells to visualize their aggregation in the cells. We found that the FL-BOULE-EGFP protein phase-separated into droplets rapidly in fewer than 30 minutes after 42  $^{\circ}$ C heat shock. Yet, Delc-BOULE-EGFP only formed some droplets 60 minutes after heat shock treatment (**Fig. 5C**).

The FRAP method can be generally used to determine the dynamics and mobility of phase-



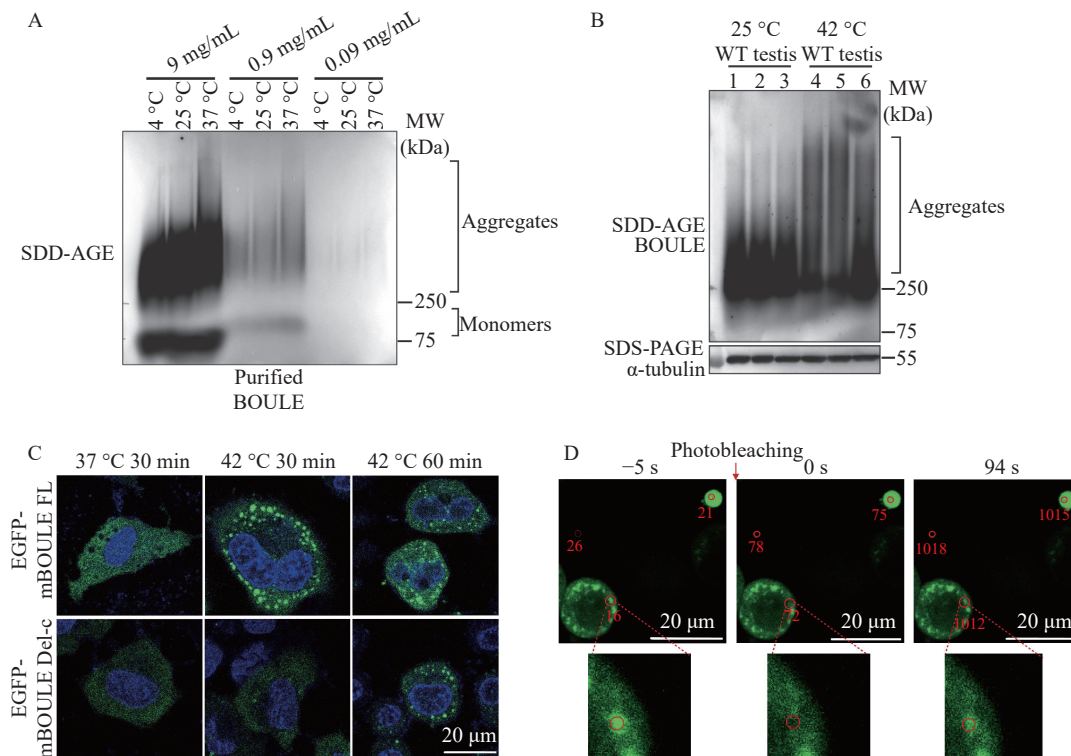
**Fig. 4 A 30-amino acid region downstream DAZ repeat is critical for BOULE protein aggregation.** A: Diagram of BOULE deletion fragments used. The blue line is the intrinsically disordered regions (IDRs) identified *via* MobiDB (<https://mobidb.bio.unipd.it/>). B: pCMV6-mouse *Boule* deletion plasmids (Del-a to Del-f) as well as pCMV6 mouse BOULE FL plasmid were transfected into HEK293T cells. Plasmids pCMV6-mouse BOULE deletion 5 (Del-5) and pCMV6-vector were used as negative controls. After 48 hours, cells were harvested for SDD-AGE and SDS-PAGE assays for BOULE aggregates and monomers by anti-BOULE antibody and anti- $\beta$ -actin antibody. RRM: RNA Recognition Motif; DAZ: Deleted in Azoospermia.

separated liquid droplet or condensate in live cells. Moreover, if EGFP is fused to a protein, the mobility of the protein of interest can be calculated following the EGFP signal in the bleached area. We determined the mobility of FL-BOULE-EGFP liquid droplets in HeLa cells by FRAP and found that the BOULE droplets exhibited a certain degree of recovery properties. This further suggests that the formation of BOULE droplets arises through phase separation (**Fig. 5D**).

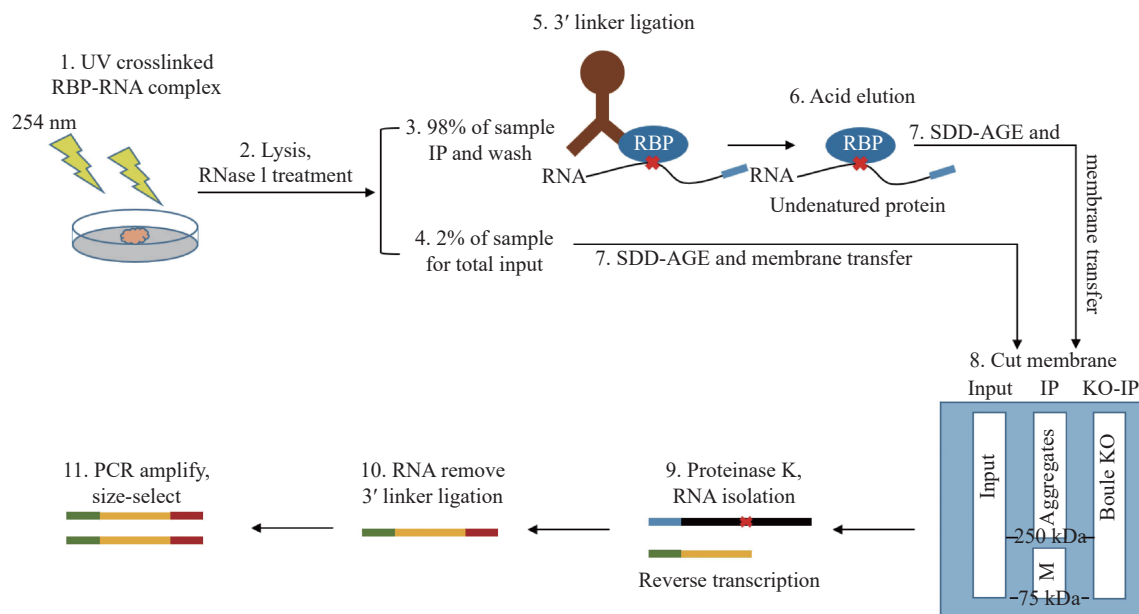
#### Identification of BOULE aggregates targets in mouse testes by eCLIP

To shed light on how BOULE aggregates function during spermatogenesis, we performed an eCLIP protocol to identify transcriptome-wide RNA targets of the BOULE aggregates (**Fig. 6**). While eCLIP has been widely used in identifying RNA targets bound by RNA binding proteins, there is no report on using eCLIP to identify mRNAs bound by RNA binding

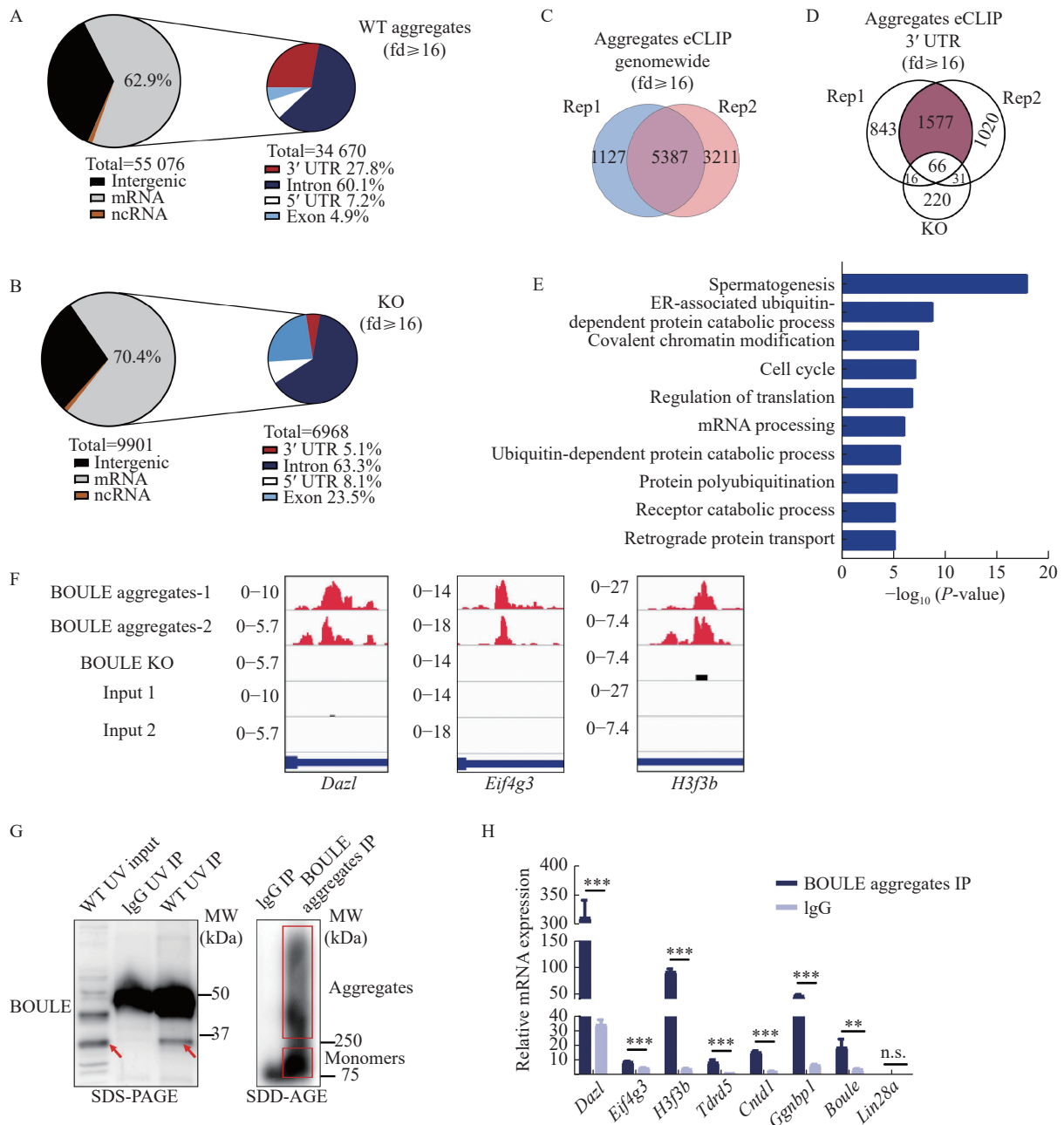
protein aggregates<sup>[28]</sup>. Mouse testes were subjected to UV-mediated crosslinking, lysis, and treatment with a limited amount of RNase, followed by immunoprecipitation (IP) of protein-RNA complexes similar to standard eCLIP procedure. To avoid the influence of targets bound by the BOULE monomer, we excised the bands above 250 kDa to avoid any overlapping targets with monomer BOULE (**Fig. 6** and **Fig. 7G**). RNA targets from two replicates from BOULE aggregates eCLIP overlapped extensively supporting the reproducibility of the targets-identification. The eCLIP peaks of at least 16-fold enrichment over the input lane were considered as BOULE aggregate targets. The resulting eCLIP transcripts from two biological replicates were filtered and mapped individually and were then intersected to identify 5387 overlapping eCLIP genomic transcripts (**Fig. 7C**). Compared with the knock-out group, we find that the 3' UTR of BOULE aggregates targets have the greatest change in the whole genome (**Fig. 7A**



**Fig. 5 Effect of temperature on BOULE protein aggregation using purified BOULE protein, EGFP tagged BOULE expressing HeLa cells, and mouse testes.** A: Concentrated BOULE solution was serially diluted into 9 mg/mL (280 mmol/L), 0.9 mg/mL (28 mmol/L), 0.09 mg/mL (2.8 mmol/L) and placed at 4 °C, 25 °C, 37 °C for 30 minutes, and then analyzed by SDD-AGE assay. B: Comparison of BOULE aggregate formation from the testes undergoing heat-shock treatment (42 °C) and from untreated testes (25 °C) by SDD-AGE assay. BOULE protein detected at molecular weight bigger than 250 KDa represented BOULE aggregates. Three replicates for each group were used. Alpha-tubulin was used as a loading control. C: Plasmids pCDNA3.1-EGFP-mBOULE FL and pCDNA3.1-EGFP-mBOULE Del-c (BOULE protein 210–240 amino acids deletion construct) were transfected into HeLa cells, after 24 hours, cells were placed at 37 °C and 42 °C for 30 minutes. Droplet formation was observed during fluorescence imaging. D: BOULE liquid droplet property was tested by examining the mobility of EGFP-mBOULE liquid droplets in HeLa cells using fluorescence recovery after photobleaching method. The recovery property of BOULE liquid droplet was shown with images of droplets before photobleaching (–5 seconds), at the time of photobleaching (0 seconds), and 94 seconds post-photobleaching. The lower panel showed blow-up images showing the recovery of droplets. Scale bars: 20  $\mu$ m.



**Fig. 6 Schematic diagram of eCLIP experiment procedure.** BOULE aggregates enhanced UV-crosslinking and immunoprecipitation (eCLIP) started with UV crosslinking of seminiferous tubules from mouse testes on postnatal day 25, followed by partial RNase digestion and immunoprecipitation. RNA-protein extracts from Input, BOULE-IP from wild type testes (IP) and *Boule* knockout testes (KO-IP) were eluded by acidic elution buffer and run by SDD-AGE, then transferred to nitrocellulose filter membrane, specific regions (250 kDa or bigger bands) were excised from BOULE IP, the membrane from 75 to 250 kDa was represented for monomer (M). Input lane and *Boule* KO lane would be excised from 75 kDa or higher shown as in step 8 of the procedure. Next, the excised protein-bound RNAs were treated with proteinase K to extract aggregate-bound RNAs. Library construction is the same as reported<sup>[20]</sup>.



**Fig. 7 Identification of BOULE aggregates targets in mouse testes by eCLIP.** A: Distribution of BOULE aggregates eCLIP peaks in different genomic regions and different parts of genic regions. B: Distribution of BOULE KO eCLIP peaks in different genic regions. C: Genome-wide overlap of aggregate target transcripts from two biological replicates of BOULE (fold enrichment ≥ 16). D: Overlap of BOULE aggregates target transcripts in 3' UTR. (fold enrichment 16). E: GO analysis result of 1577 BOULE aggregates targets. F: Integrative genomics viewer track of three BOULE aggregates binding targets, *Dazl*, *Eif4g3*, and *H3f3b*, were presented, with peaks from two biological replicates of BOULE IP, one BOULE KO and two independent Inputs. The peaks are only located on 3' UTRs (the narrow horizontal bar). G and H: Validation of BOULE aggregates eCLIP results by selected target RNA IP. Relative mRNA expression of each target was calculated based on the ratio of their expression in the IP relative to IgG using *Gapdh* as the internal control. Non-target, *Lin28a*, was not enriched. Data are expressed as mean±SD. Statistical analyses were performed by *t*-test for two-group comparisons. \*\**P*<0.01; \*\*\**P*<0.001. n.s.: no significance.

and **B**). We thus focused our analysis on mRNA targets with BOULE aggregates binding sites at the 3' UTR region. After removing the influence of *Boule* knockout and monomer non-specific targets, we found that BOULE aggregates bind to 1577 transcripts in adult testis on their 3' UTR (**Fig. 7D**). Those targets

were significantly enriched in spermatogenesis pathways (classified by DAVID). Please see **Fig. 7E** for further details.

We further confirmed the quality and reproducibility of the two libraries by comparing the binding sites and peaks among the targets by Genome



browser tracks (<http://genome.ucsc.edu/>). We observed a consistent peak distribution for the same target transcripts among the two libraries (**Fig. 7F**). To determine the validity of the BOULE aggregates target mRNAs identified by eCLIP, we randomly selected six targets from spermatogenesis pathways among the 1577 3' UTR targets shared by two libraries and found they were all significantly enriched in BOULE aggregates IP. *Boule* is used as positive control, by contrast, non-targets such as *Lin28a* were not enriched (**Fig. 7G and H**). This suggested that the 1577 shared targets were reliable BOULE aggregates targets. Hence, we conclude that the BOULE aggregates specifically bind to a large number of transcripts, leading us to propose that BOULE aggregates may have a physiological function during sperm development.

## Discussion

In response to stimulus, cells constantly face the demanding task of promptly organizing various molecular reactions. To address this challenge, cells have evolved compartments which facilitate spatiotemporal control of biological reactions. Many membrane-less organelles, such as stress granules and P-bodies, dynamically assemble and disassemble in response to cellular stress and other stimuli<sup>[29–30]</sup>. A number of studies have identified many RNA-binding proteins, such as hnRNPA1, and TDP43, which contribute to this conditional cellular compartmentalization. These proteins commonly contain LCDs or IDRs that can drive full-length proteins into LLPS through weak multivalent interactions<sup>[25–26]</sup>. However, at high concentrations, many proteins form amyloid-like filaments consisting of repeated cross-beta strand elements<sup>[31–32]</sup>, which typically have been viewed as pathological entities associated with neurodegenerative disease and prionopathies<sup>[33]</sup>. Although, evidence increasingly indicates that most proteins can adopt the functional amyloid fold under certain conditions and that some amyloids carry-out non-pathologic biological roles.

When proteins aggregate, they usually form into an amyloid state consisting of cross  $\beta$ -sheet structure forming long fibrils. At the same time, amyloids also exhibit several other unusual properties, such as detergent and protease resistance, interaction with specific dyes, and the ability to induce the transition of some proteins from a soluble form to an aggregated one. These same features have also been demonstrated for the aggregates lacking cross- $\beta$  structure, which are commonly called "amyloid-like". The conformational

switch into an amyloid-like state is often mediated by multivalent interactions among proteins containing low-complexity regions, which are intrinsically disordered.

In this study, we found that the testicular-specific BOULE protein can form amyloid-like aggregation *in vitro* and *in vivo*, and BOULE aggregates showed identical developmental expression patterns (detected by SDD-AGE) with BOULE monomers (detected by SDS-PAGE). Through different BOULE protein deletion experiments, we found that the key region essential for the formation of BOULE aggregation was located with the peptide of #210-240AA, which happens to be located in the IDR region. The aggregation of BOULE was affected by temperature, when temperature increased the aggregation of BOULE protein increased significantly and rapidly. Interestingly, testicular BOULE proteins also undergo phase separation after heat shock, and FL-BOULE-EGFP droplets exhibit clear mobility after laser treatment, supporting BOULE protein undergoes aggregation. Phase transition may be one of the transient states during BOULE aggregation into amyloid-like state. As RNA binding proteins function through binding RNA targets, we are asking whether BOULE aggregates might bind to any RNAs. We also found that BOULE aggregates bind to a large number of spermatogenesis-related mRNAs by eCLIP, supporting potential functional roles of BOULE amyloid-like aggregates.

Amyloids are fibrous protein aggregates that are known for their roles in the etiology of neurodegenerative diseases. The formation of amyloid is a multi-step process in which proteins progress from soluble oligomers to fibrils. Proteins which form functional amyloids follow similar aggregation pathways to those that form pathological amyloids, including the development of cytotoxic intermediate oligomeric amyloids<sup>[34]</sup>. Of the functional amyloids, many proteins transition quite rapidly under controlled cellular conditions from monomeric to mature amyloid forms, possibly as a means to avoid cytotoxic intermediates<sup>[3]</sup>. Indeed, unlike pathological amyloids such as  $\alpha$ -synuclein, which require days to weeks to form mature amyloid fibrils *in vitro*, proteins forming functional amyloids often do so within minutes to hours. Our results showed that the kinetics of BOULE amyloidogenesis is similar to consistent with other proteins that form functional amyloids by showing rapid rates of amyloid formation *in vitro*. Other reproductive functional amyloids also exhibit fast rates of the amyloid formation including Xvelo, the amyloid component of Balbiani bodies, which transitioned into a fibrillar matrix after 12–24 hours<sup>[4–35]</sup>. Similarly, fibrils were detected in recombinant Rim4 after overnight incubation<sup>[1]</sup>.

Previous studies have shown that budding yeast builds massive RNA-binding protein structures that exhibit biochemical properties of amyloid to regulate translation during meiosis. Reported existence of amyloid-like assemblies in mouse and frog meiosis led to the hypothesis that amyloid-like assemblies are a conserved feature of gametogenesis<sup>[1]</sup>. Our work on BOULE aggregate formation in the testis supports the evolutionary conservation of functional amyloid-like RNA binding proteins during mammalian spermatogenesis. Further study on DAZ family proteins' aggregate formation and their RNA target expression may elaborate our understanding of the function of the amyloid-like states in DAZ proteins. Establishing mouse models disrupting aggregation without impacting protein expression of BOULE will be a key to delineating the function of BOULE or DAZL amyloid-like aggregates on spermatogenesis. In combination with eCLIP target comparisons between aggregates and monomers, we should also investigate the molecular mechanisms involved in BOULE protein amyloid-like aggregates function to promote sperm development. We also need clues about what creates the conservation of DAZ family proteins in metazoan gametogenesis.

### Acknowledgments

We are grateful for the helpful discussion with Dr. Gail Cornwall throughout the course of this project. We would like to thank Dr. Liming Li for comments on our manuscript. This work is supported by the National Natural Science Foundation of China (No. 3197060323), SKLRM grant (SKLRM-2019B2) and the Jiangsu ShuangChuang Talent Program, as well as the Jiangsu graduate student innovation fellowship to Y.S.

### References

- [1] Berchowitz LE, Kabachinski G, Walker MR, et al. Regulated formation of an amyloid-like translational repressor governs gametogenesis[J]. *Cell*, 2015, 163(2): 406–418.
- [2] Carpenter K, Bell RB, Yunus J, et al. Phosphorylation-mediated clearance of amyloid-like assemblies in meiosis[J]. *Dev Cell*, 2018, 45(3): 392–405.e6.
- [3] Fowler DM, Koulov AV, Alory-Jost C, et al. Functional amyloid formation within mammalian tissue[J]. *PLoS Biol*, 2006, 4(1): e6.
- [4] Hewetson A, Do HQ, Myers C, et al. Functional amyloids in reproduction[J]. *Biomolecules*, 2017, 7(3): 46.
- [5] Hewetson A, Khan NH, Dominguez MJ, et al. Maturation of the functional mouse CRES amyloid from globular form[J]. *Proc Natl Acad Sci U S A*, 2020, 117(28): 16363–16372.
- [6] Nott TJ, Petsalaki E, Farber P, et al. Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles[J]. *Mol Cell*, 2015, 57(5): 936–947.
- [7] Shin Y, Brangwynne CP. Liquid phase condensation in cell physiology and disease[J]. *Science*, 2017, 357(6357): eaaf4382.
- [8] Matzuk MM, Burns KH. Genetics of mammalian reproduction: modeling the end of the germline[J]. *Annu Rev Physiol*, 2012, 74: 503–528.
- [9] Alberti S, Halfmann R, King O, et al. A systematic survey identifies prions and illuminates sequence features of prionogenic proteins[J]. *Cell*, 2009, 137(1): 146–158.
- [10] Shah C, VanGompel MJM, Naeem V, et al. Widespread presence of human *BOULE* homologs among animals and conservation of their ancient reproductive function[J]. *PLoS Genet*, 2010, 6(7): e1001022.
- [11] Xu EY, Moore FL, Pera RAR. A gene family required for human germ cell development evolved from an ancient meiotic gene conserved in metazoans[J]. *Proc Natl Acad Sci U S A*, 2001, 98(13): 7414–7419.
- [12] VanGompel MJW, Xu EY. The roles of the DAZ family in spermatogenesis: more than just translation?[J]. *Spermatogenesis*, 2011, 1(1): 36–46.
- [13] Couthouis J, Hart MP, Shorter J, et al. A yeast functional screen predicts new candidate ALS disease genes[J]. *Proc Natl Acad Sci U S A*, 2011, 108(52): 20881–20890.
- [14] King OD, Gitler AD, Shorter J. The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease[J]. *Brain Res*, 2012, 1462: 61–80.
- [15] Xu EY, Lee DF, Klebes A, et al. Human *BOULE* gene rescues meiotic defects in infertile flies[J]. *Hum Mol Genet*, 2003, 12(2): 169–175.
- [16] VanGompel MJW, Xu EY. A novel requirement in mammalian spermatid differentiation for the DAZ-family protein Boule[J]. *Hum Mol Genet*, 2010, 19(12): 2360–2369.
- [17] Bellvé AR. Purification, culture, and fractionation of spermatogenic cells[J]. *Methods Enzymol*, 1993, 225: 84–113.
- [18] Halfmann R, Lindquist S. Screening for amyloid aggregation by semi-denaturing detergent-agarose gel electrophoresis[J]. *J Vis Exp*, 2008, (17): 838.
- [19] Ji S, Luo Y, Cai Q, et al. LC domain-mediated coalescence is essential for otu enzymatic activity to extend *Drosophila* lifespan[J]. *Mol Cell*, 2019, 74(2): 363–377.e5.
- [20] Van Nostrand EL, Pratt GA, Shishkin AA, et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP)[J]. *Nat Methods*, 2016, 13(6): 508–514.
- [21] Bagriantsev SN, Kushnirov VV, Liebman SW. Analysis of amyloid aggregates using agarose gel electrophoresis[J]. *Methods Enzymol*, 2006, 412: 33–48.
- [22] Berchowitz LE, Gajadhar AS, van Werven FJ, et al. A developmentally regulated translational control pathway establishes the meiotic chromosome segregation pattern[J]. *Genes Dev*, 2013, 27(19): 2147–2163.

- [23] Adiutori R, Aarum J, Zubiri I, et al. The proteome of neurofilament-containing protein aggregates in blood[J]. *Biochem Biophys Rep*, 2018, 14: 168–177.
- [24] Pani I, Madhu P, Najiya N, et al. Differentiating conformationally distinct Alzheimer's amyloid- $\beta$  oligomers using liquid crystals[J]. *J Phys Chem Lett*, 2020, 11(21): 9012–9018.
- [25] Molliex A, Temirov J, Lee J, et al. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization[J]. *Cell*, 2015, 163(1): 123–133.
- [26] Pak CW, Kosno M, Holehouse AS, et al. Sequence determinants of intracellular phase separation by complex coacervation of a disordered protein[J]. *Mol Cell*, 2016, 63(1): 72–85.
- [27] Kim B, Cooke HJ, Rhee K. DAZL is essential for stress granule formation implicated in germ cell survival upon heat stress[J]. *Development*, 2012, 139(3): 568–578.
- [28] Van Nostrand EL, Freese P, Pratt GA, et al. A large-scale binding and functional map of human RNA-binding proteins[J]. *Nature*, 2020, 583(7818): 711–719.
- [29] Protter DSW, Parker R. Principles and properties of stress granules[J]. *Trends Cell Biol*, 2016, 26(9): 668–679.
- [30] Youn JY, Dyakov BJA, Zhang J, et al. Properties of stress granule and P-body proteomes[J]. *Mol Cell*, 2019, 76(2): 286–294.
- [31] Lin Y, Protter DSW, Rosen MK, et al. Formation and maturation of phase-separated liquid droplets by RNA-binding proteins[J]. *Mol Cell*, 2015, 60(2): 208–219.
- [32] Kato M, Han TW, Xie S, et al. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels[J]. *Cell*, 2012, 149(4): 753–767.
- [33] Vaquer-Alicea J, Diamond MI. Propagation of protein aggregation in neurodegenerative diseases[J]. *Annu Rev Biochem*, 2019, 88: 785–810.
- [34] Kaye R, Head E, Thompson JL, et al. Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis[J]. *Science*, 2003, 300(5618): 486–489.
- [35] Boke E, Ruer M, Wühr M, et al. Amyloid-like self-assembly of a cellular compartment[J]. *Cell*, 2016, 166(3): 637–650.

Submit to the *Journal* by ScholarOne Manuscripts at  
<http://mc03.manuscriptcentral.com/jbrint>

