

RESEARCH NOTE

Open Access



Degenerate codon mixing for PCR-based manipulation of highly repetitive sequences

Dhanushika Ratnayake^{1,2}, Morgan Newman¹ and Michael Lardelli^{1*}

Abstract

Objective: Repeat expansion of polyglutamine tracks leads to a group of inherited human neurodegenerative disorders. Studying such repetitive sequences is required to gain insight into the pathophysiology of these diseases. PCR-based manipulation of repetitive sequences, however, is challenging due to the absence of unique primer binding sites or the generation of non-specific products.

Results: We have utilised the degeneracy of the genetic code to generate a polyglutamine sequence with low repeat similarity. This strategy allowed us to use conventional PCR to generate multiple constructs with approximately defined numbers of glutamine repeats. We then used these constructs to measure the in vivo variation in autophagic degradation activity related to the different numbers of glutamine repeats, providing an example of their applicability to study repeat expansion diseases. Our simple and easily generalised method of generating low repetition DNA sequences coding for uniform stretches of amino acid residues provides a strategy for generating particular lengths of polyglutamine tracts using standard PCR and cloning protocols.

Keywords: Polyglutamine repeats, Codon redundancy, Degeneracy of the genetic code, PCR based repeat region amplification, Aggregate proteins, Autophagy

Introduction

The aberrant expansion of unstable CAG repeats coding for polyglutamine (polyQ) tracts underlies a group of neurodegenerative diseases, including Huntington's disease (HD) and several forms of spino-cerebellar ataxia (SCA) [1]. These diseases exhibit polyQ length-dependent toxicity, whereby age at disease onset is inversely correlated to the number of polyQ repeats [2]. They display common cellular and molecular mechanisms including protein aggregation and inclusion body formation [3]. Such protein aggregates depend strongly on autophagy for their clearance and dysfunction of this pathway may contribute to the pathology of these diseases [4]. Enhancement of autophagy has been suggested to have possible therapeutic value in such diseases showing protein aggregation by promoting the clearance of these aggregates and protecting cells

against their toxic effects [5, 6]. However, studying the influence of polyQ tract length on aggregation kinetics is challenging due to difficulties faced when cloning repetitive DNA sequences primarily due to the lack of unique primer binding sites [7, 8]. Previously, several polymerase chain reaction (PCR) based methods to amplify repetitive DNA regions have been described [9–11]. However, most of these either generate nonspecific products, flawed repeats, or a collection of clones with varying numbers of repeats making the identification and isolation of the specific clone of interest laborious [12–14]. In order to investigate the autophagic degradation activity or 'autophagic flux' of polyQ protein aggregates we sought to clone reporter constructs containing more closely defined numbers of glutamine residues. We designed a polyQ sequence with low repeat similarity by exploiting the codon redundancy of the genetic code. This strategy allowed us to amplify close to the desired numbers of glutamine repeats (although still with some variability due to two distinct causes), which we subsequently used to assess in vivo

*Correspondence: michael.lardelli@adelaide.edu.au

¹ School of Biological Sciences, Alzheimer's Disease Genetics Laboratory, University of Adelaide, North Terrace, Adelaide, SA 5005, Australia
Full list of author information is available at the end of the article

variations in ‘autophagic flux’ in a larval zebrafish model for Alzheimer’s disease [15].

Main text

Results and discussion

The ‘autophagic flux’ assay described in Jiang et al. [15] is a quantitative green fluorescent protein (GFP) reporter assay that measures the ratiometric changes of polyQ-GFP to free GFP via Western blot analysis. A multicistronic reporter construct was designed to code for two proteins; polyQ linked to N-terminal GFP and free GFP. The viral 2A (v2A) sequence was placed as a linker region between the sequences coding for the two proteins to enable stoichiometric translation of two separate proteins from one open reading frame (Fig. 1a) [16].

We designed a polyQ sequence containing 80 glutamine repeats (Q_{80}). The sequence was designed to have low repeat similarity by randomly interspacing glutamine-coding CAG triplets with glutamine-coding CAA triplets (Fig. 1b, c). The nucleotide substitutions were made by eye to generate a semi random pattern. This non-repetitive sequence design should not only enhance sequence stability during propagation in bacteria but also enabled the design of PCR primers that annealed to specific regions of the sequence.

The Q_{80} -GFP-v2A-GFP construct described above was commercially synthesised (Biomatik Corporation) (see Additional file 1) and sub-cloned via the *Bam*HI and *Cl*aI restriction sites into the Tol2 transposon-based, pT2AL200R150G gene transfer vector (hereafter referred to as Tol2) available from the Kawakami laboratory [17] (Fig. 1a and see Additional file 2).

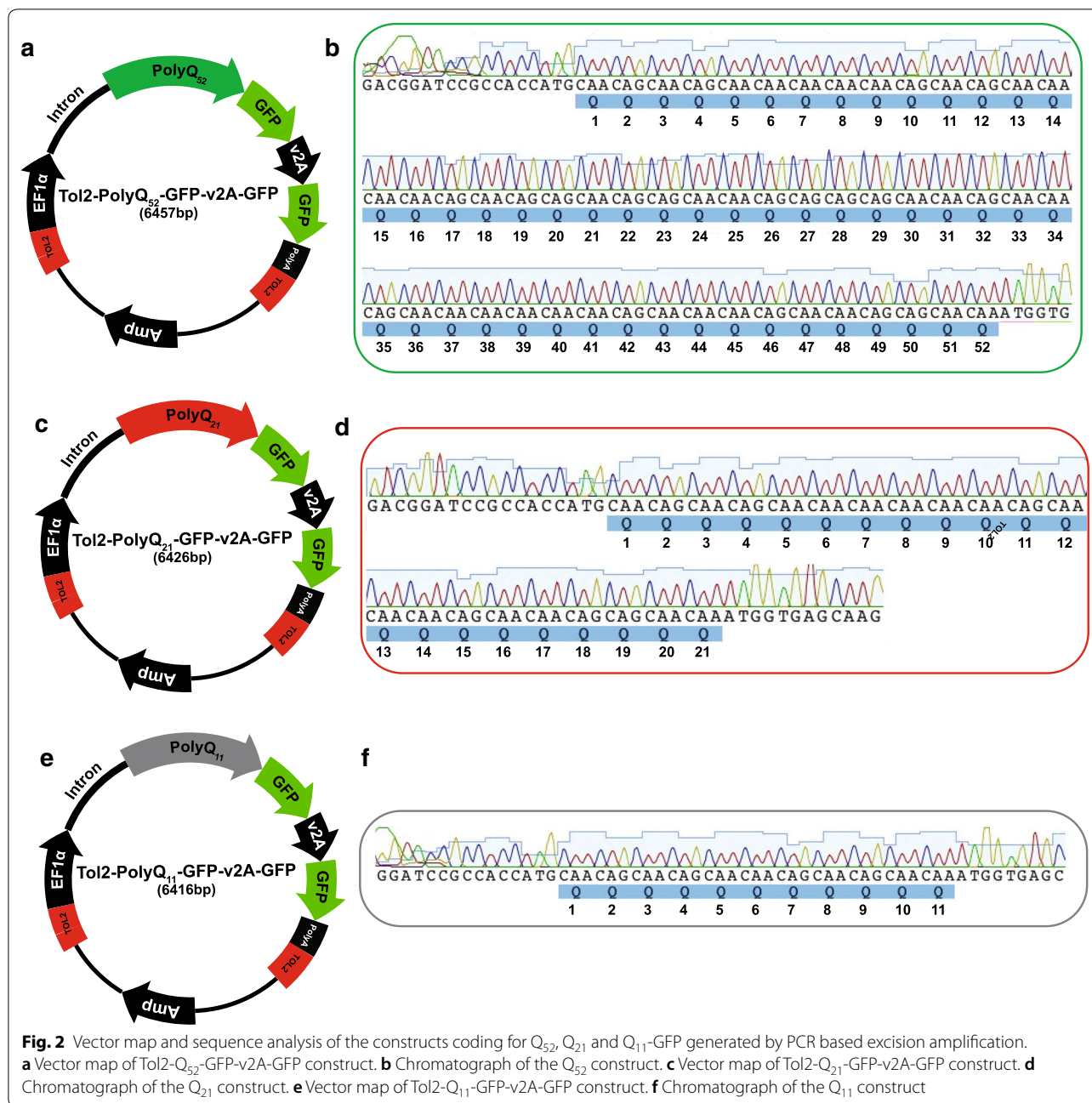
PolyQ constructs with lower numbers of glutamine repeats were generated by PCR-based exclusion amplification of the Tol2- Q_{80} -GFP-v2A-GFP construct. The primers were designed to amplify around the vector excluding a defined number of glutamine repeats to generate the constructs of interest (Fig. 1d–f). We aimed to generate vectors with approximately 52 (Q_{52}), 31 (Q_{31}) and 10 (Q_{10}) glutamine repeats. The putative Q_{52} and Q_{31} vectors were generated using the same reverse primer coupled with different forward primers. This common reverse primer amplified 2 glutamine repeats, while the forward primers amplified the additional 50 and 29

glutamine repeats needed to generate the Q_{52} and Q_{31} vectors, respectively. The position of the reverse primer was shifted slightly to optimise the amplification of the putative Q_{10} vector, such that the reverse primer now amplified 5 glutamine repeats while the forward primer amplified the remaining 5 glutamine repeats (Fig. 1d–f). By using stringent annealing temperatures in the PCR reaction we obtained specific primer binding. Gel extracted and purified PCR products were phosphorylated, circularised by self-ligation and subsequently transformed into competent cells (see Additional file 3). PCR using primers that flanked the polyQ region showed approximately expected product sizes for the intended putative Q_{52} , Q_{31} and Q_{10} vectors (Fig. 1g). The generated constructs were sequenced to determine whether the expected polyQ repeat numbers were present. While the Q_{52} construct had the expected number of glutamine repeats (Fig. 2a, b), sequencing revealed minor discrepancies with expected polyQ numbers for the other two constructs, where the Tol2- Q_{31} -GFP-v2A-GFP vector had 21-glutamine repeats (hereafter referred to as Tol2- Q_{21} -GFP-v2A-GFP) (Fig. 2c, d) and the Tol2- Q_{10} -GFP-v2A-GFP vector had 11-glutamine repeats (hereafter referred to as Tol2- Q_{11} -GFP-v2A-GFP) (Fig. 2e, f). Further analysis revealed that the generated Q_{21} sequence was derived directly from the original sequence and the loss of glutamine repeats was due to the Q_{31} forward primer binding 30 bp downstream from the predicted binding site. In contrast, the additional glutamine repeat in the Q_{11} sequence was generated de novo, an addition of a CAA codon.

To study the aggregation kinetics and ‘autophagic flux’ of polyQ protein in vivo, we injected the generated polyQ vectors (25 ng/ μ L) and transposase mRNA (25 ng/ μ L) into groups of one-cell-stage zebrafish embryos (Fig. 3a, b). Western blot analysis with anti-GFP antibody of 24 h post fertilisation (hpf) embryo lysates (10 embryos per sample) was carried out for each group. The empty Tol2 vector (25 ng/ μ L) and transposase mRNA (25 ng/ μ L) injected and un-injected embryos were included as controls. Embryos injected with the polyQ constructs produced two bands detected by the anti-GFP antibody as expected; GFP attached to polyQ (poly Q_{80} -GFP at ~48 kDa, poly Q_{52} -GFP at ~38 kDa, poly Q_{21} -GFP at ~33 kDa and poly Q_{11} -GFP

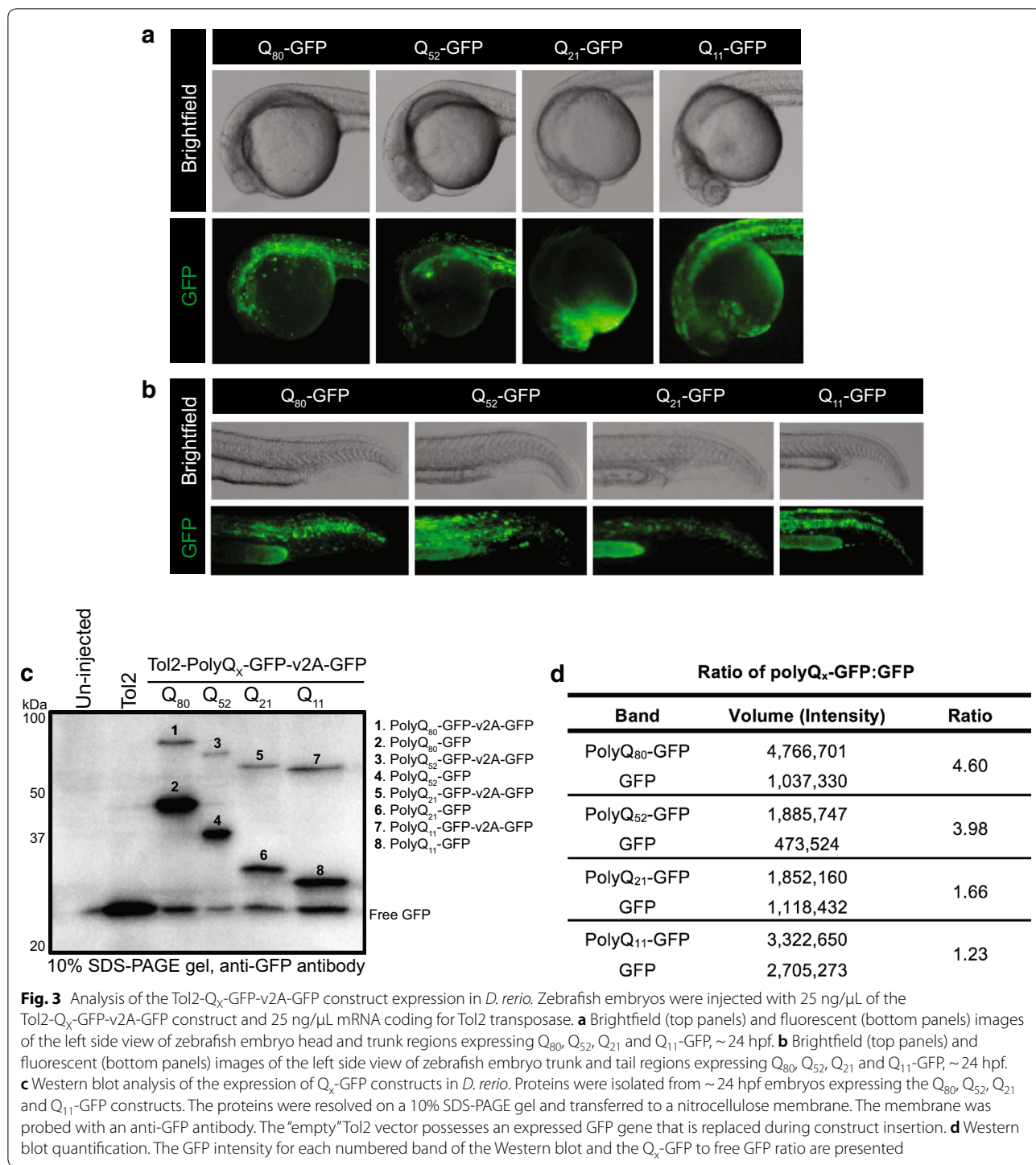
(See figure on next page.)

Fig. 1 Generating the putative Q_{52} , Q_{31} and Q_{10} -GFP constructs using the Tol2- Q_{80} -GFP-v2A-GFP multicistronic reporter construct. **a** Vector map of the Tol2- Q_{80} -GFP-v2A-GFP construct. **b** Summary of CAA and CAG degenerate codon usage in this construct. **c** Chromatograph of the construct depicting the randomly interspaced glutamine coding CAG and CAA triplets. **d** Schematic illustration of PCR-based exclusion amplification of the Q_{80} -GFP containing construct to generate polyQ constructs with lower numbers of glutamine repeats. **e** Q_{80} sequence showing Q_{52} , Q_{31} and Q_{10} primer binding sites. **f** Sequences of primers intended to generate Q_{52} , Q_{31} and Q_{10} constructs. **g** Analytical agarose gel electrophoresis for each of the putative Q_{80} , Q_{52} , Q_{31} and Q_{10} vectors using primers flanking the polyQ region. PCR product sizes of ~270, ~180, ~120 and ~60 bp for the intended putative Q_{80} , Q_{52} , Q_{31} and Q_{10} -GFP constructs, respectively, were seen



at ~31 kDa), and free GFP (~27 kDa) (Fig. 3c). Each of the polyQ-GFP construct expressing embryo lysates also showed a fainter band of higher protein size corresponding to the full length polyQ_x-GFP-v2A-GFP construct (polyQ₈₀-GFP-v2A-GFP at ~75 kDa, polyQ₅₂-GFP-v2A-GFP at ~65 kDa, polyQ₂₁-GFP-v2A-GFP at ~60 kDa, and polyQ₁₁-GFP-v2A-GFP at ~58 kDa). This band represents the ~10% of the total protein that is translated as a single, full-length protein when using the v2A sequence system [18]. The Q₈₀-GFP:GFP, Q₅₂-GFP:GFP, Q₂₁-GFP:GFP and

Q₁₁-GFP:GFP ratios are ~5, ~4, ~2 and ~1, respectively (Fig. 3d). As the v2A sequence allows for the stoichiometric translation of the polyQ-GFP and GFP proteins, in theory the polyQ-GFP:GFP ratio should be 1. The greater ratios observed may indicate an accumulation of those proteins. These observations are in agreement with the literature, where it has been shown that polyQ-GFP fusion constructs containing greater than 19 glutamine residues aggregate within transfected cells in a length-dependent manner [19]. These observations lead us to conclude that our



Tol2-Q_x-GFP-v2A-GFP constructs provide a useful tool to study ‘autophagic flux’ in vivo in a larval zebrafish model.

Conclusion

In conclusion this study provides a robust and easily adoptable solution to generate close to intended lengths

of polyQ repeats. In order to generate exact numbers of glutamine repeats subsequent rounds of amplifications with altered primer sequences can be carried out. In addition the primer length can be increased to enhance specificity of primer annealing to the template. Our technique has several advantages over the existing methods

for PCR-based amplification of repetitive regions and aims to minimise the generation of non-specific PCR products and flawed repeats by exploiting the codon redundancy of the genetic code to generate a synonymous DNA coding sequence with reduced repetition. In addition, our approach is not limited to generating polyQ repeat sequences but can also be generalised to generate other nucleotide repeat sequences. Furthermore, our method is relatively cheap, as only the initial polyQ₈₀-GFP-v2A-GFP construct requires commercial synthesis, the cost of which depends on the price per nucleotide base, length, purity and mass. All other materials required are standard reagents used for molecular cloning. In summary the technique described provides an easy to adopt, affordable solution to generate repeat-coding DNA sequences that can be manipulated as required.

Limitations

- Initial construct requires commercial synthesis.
- May not generate the exact number of glutamine repeats required.

Additional files

Additional file 1. Sequence design for the Q₈₀-GFP-v2A-GFP construct. The commercially synthesised Q₈₀-GFP-v2A-GFP construct is flanked by *Bam*HI I and *Cla*I I restriction sites used for sub cloning into the desired final vector.

Additional file 2. Sub-cloning the Q₈₀-GFP-v2A-GFP construct into the Tol2 vector. The Q₈₀-GFP-v2A-GFP construct provided in the pBluescript II SK(+) vector is sub-cloned into the pT2AL200R150G (Tol2) vector.

Additional file 3. Exclusion amplification of the Tol2-Q₈₀-GFP-v2A-GFP construct to generate constructs containing putative Q₅₂, Q₃₁ or Q₁₀ repeats. Detailed method for generating polyQ constructs with lower number of glutamine repeats using the Tol2-Q₈₀-GFP-v2A-GFP construct as a template.

Abbreviations

Q₈₀: 80 glutamine repeats; GFP: green fluorescent protein; hpf: hours post fertilisation; HD: Huntington's disease; polyQ: polyglutamine; PCR: polymerase chain reaction; SCA: spino-cerebellar ataxia; v2A: viral 2A.

Authors' contributions

DR performed the experimental work and drafted this manuscript. MN supervised the laboratory work and trained DR in the necessary techniques. ML conceived the project and edited the manuscript. All authors read and approved the final manuscript.

Author details

¹ School of Biological Sciences, Alzheimer's Disease Genetics Laboratory, University of Adelaide, North Terrace, Adelaide, SA 5005, Australia. ² Present Address: Australian Regenerative Medicine Institute, Monash University, Wellington Road, Clayton 3800, Australia.

Acknowledgements

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

All data generated or analysed during this study are included in this published article and its additional information files.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Conducted under the auspices of the University of Adelaide Animal Ethics Committee under permit S-2014-108.

Funding

This work was supported by a Grant from Australia's National Health and Medical Research Council, GNT1061006.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 17 January 2018 Accepted: 20 March 2018

Published online: 27 March 2018

References

- Orr HT. Beyond the Qs in the polyglutamine diseases. *Genes Dev.* 2001;15(8):925–32.
- Perutz MF, Windle A. Cause of neural death in neurodegenerative diseases attributable to expansion of glutamine repeats. *Nature.* 2001;412(6843):143–4.
- Ross CA, Poirier MA. Protein aggregation and neurodegenerative disease. *Nat Med.* 2004;10(7):S10–7.
- Rubinsztein DC. The roles of intracellular protein-degradation pathways in neurodegeneration. *Nature.* 2006;443(7113):780–6.
- Fleming A, Noda T, Yoshimori T, Rubinsztein DC. Chemical modulators of autophagy as biological probes and potential therapeutics. *Nat Chem Biol.* 2011;7(1):9.
- Ravikumar B, Vacher C, Berger Z, Davies JE, Luo S, Oroz LG, Scaravilli F, Easton DF, Duden R, O'Kane CJ. Inhibition of mTOR induces autophagy and reduces toxicity of polyglutamine expansions in fly and mouse models of Huntington disease. *Nat Genet.* 2004;36(6):585.
- Samadashwily GM, Raca G, Mirkin SM. Trinucleotide repeats affect DNA replication in vivo. *Nat Genet.* 1997;17(3):298.
- Godiska R, Mead D, Dhodda V, Wu C, Hochstein R, Karsi A, Usdin K, Entezam A, Ravin N. Linear plasmid vector for cloning of repetitive or unstable sequences in *Escherichia coli*. *Nucleic Acids Res.* 2009;38(6):e88.
- Riet J, Ramos L, Lewis RV, Marins L. Improving the PCR protocol to amplify a repetitive DNA sequence. *Genet Mol Res.* 2017;16(3). <https://doi.org/10.4238/gmr16039796>.
- Sahdev S, Saini S, Tiwari P, Saxena S, Saini KS. Amplification of GC-rich genes by following a combination strategy of primer design, enhancers and modified PCR cycle conditions. *Mol Cell Probes.* 2007;21(4):303–7.
- Hommelsheim CM, Frantzeskakis L, Huang M, Ülker B. PCR amplification of repetitive DNA: a limitation to genome editing technologies and many other applications. *Sci Rep.* 2014;4:5052.
- Takahashi N, Sasagawa N, Suzuki K, Ishiura S. Synthesis of long trinucleotide repeats in vitro. *Neurosci Lett.* 1999;262(1):45–8.
- Laccone F, Maiwald R, Bingemann S. A fast polymerase chain reaction-mediated strategy for introducing repeat expansions into CAG-repeat containing genes. *Hum Mutat.* 1999;13(6):497.
- Peters MF, Ross CA. Preparation of human cDNAs encoding expanded polyglutamine repeats. *Neurosci Lett.* 1999;275(2):129–32.
- Jiang H, Newman M, Ratnayake D, Lardelli M. Ratiometric assays of autophagic flux in zebrafish for analysis of familial Alzheimer's disease-like mutations. *bioRxiv.* 2018. <https://doi.org/10.1101/272351>.
- Provost E, Rhee J, Leach SD. Viral 2A peptides allow expression of multiple proteins from a single ORF in transgenic zebrafish embryos. *Genesis.* 2007;45(10):625–9.

17. Kawakami K. Tol2: a versatile gene transfer vector in vertebrates. *Genome Biol.* 2007;8(1):S7.
18. de Felipe P, Luke GA, Hughes LE, Gani D, Halpin C, Ryan MD. E unum pluribus: multiple proteins from a self-processing polyprotein. *Trends Biotechnol.* 2006;24(2):68–75.
19. Moulder KL, Onodera O, Burke JR, Strittmatter WJ, Johnson EM. Generation of neuronal intranuclear inclusions by polyglutamine-GFP: analysis of inclusion clearance and toxicity as a function of polyglutamine length. *J Neurosci.* 1999;19(2):705–15.

Submit your next manuscript to BioMed Central
and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

