Note to copy editor : Non-standard (review) chapter format

Methods for assessing DNA repair and repeat expansion in Huntington's Disease.

Thomas Massey, Branduff McAllister and Lesley Jones*

(Cardiff University, Cardiff, UK)

* Corresponding author

Institute of Psychological Medicine and Clinical Neurosciences

MRC Centre for Neuropsychiatric Genetics and Genomics

Cardiff University

Hadyn Ellis Building

Maindy Road, Cathays

Cardiff

CF24 4HQ

Email: Jonesl1@cf.ac.uk, mcallisterw@cf.ac.uk, masseyt1@cf.ac.uk)

Running head

DNA repair and CAG expansion in HD

Abstract

Huntington's disease (HD) is caused by a CAG repeat expansion in the *HTT* gene. Repeat length can change over time, both in individual cells and between generations, and longer repeats may drive pathology. Cellular DNA repair systems have long been implicated in CAG repeat instability but recent genetic evidence from humans linking DNA repair variants to HD onset and progression has reignited interest in this area. The DNA damage response plays an essential role in maintaining genome stability, but may also licence repeat expansions in the context of HD. In this chapter we summarise the methods developed to assay CAG repeat expansion/contraction in vitro and in cells, and review the DNA repair genes tested in mouse models of HD. While none of these systems is currently ideal, new technologies, such as long-read DNA sequencing, should improve the sensitivity of assays to assess the effects of DNA repair pathways in HD. Improved assays will be essential precursors to high throughput testing of small molecules that can alter specific steps in DNA repair pathways and perhaps ameliorate expansion or enhance contraction of the *HTT* CAG repeat.

Keywords

Huntington's disease ; Neurodegeneration ; Trinucleotide repeat ; CAG repeat ; Genetic modifiers ; DNA repair ; DNA damage response ; Mouse models ; Cell models ; Somatic expansion

1. Introduction

Huntington's disease (HD) is caused by an expanded CAG repeat in exon 1 of the *HTT* gene. Wild-type alleles contain between 9 and 26 repeats, whereas disease-causing alleles contain at least 36 repeats. The longer the repeat, the earlier the onset of motor symptoms, although there is considerable variation. Intermediate alleles containing between 27 and 35 repeats do not cause disease but are a risk factor for further expansion into the pathogenic range in the subsequent generation. By definition, repeat expansion requires DNA synthesis. In dividing cells this may occur in the context of DNA replication or DNA repair, whereas in non-dividing ('post-mitotic') cells such as neurons only DNA repair pathways can be implicated. There is increasing evidence from genetics, animal models and cellular studies for the involvement of DNA repair in HD pathogenesis.

1.1. Genetic evidence for the role of DNA repair in HD

A recent genome-wide association study looking for single-nucleotide polymorphisms associated with particularly early or late motor onset of disease identified significant signals in or near DNA repair genes [1]. A pathway analysis in these data detected a strong signal in the genes of the DNA damage response implicating this pathway in modifying age at motor onset of disease. More specifically, signals were enriched for pathways involving mismatch repair genes and these findings have since been corroborated by the association of a coding variant in *MSH3* with HD progression ([1, 2]; see chapter by Stone and Holmans for more detail [3].

In theory, DNA repair could be involved in HD either at the level of the CAG repeat in the *HTT* gene or downstream of HTT protein function. Looking across the spinocerebellar ataxias, which are caused by expanded CAG repeats in various genes, highlights the same genetic signal as in HD, suggesting that there is a common pathogenic mechanism acting on CAG repeats in DNA [4, 5]. It remains possible that DNA repair is also affected by mutant HTT protein, or that an expanded polyglutamine can have effects on DNA repair outwith of protein context [6], but most evidence points towards repair processes affecting CAG repeat stability in DNA.

1.2. CAG repeat instability in HD

DNA microsatellites consist of short (up to 5 base pair (bp)) sequences that are repeated in tandem multiple times at specific genomic loci. Microsatellites are prone to expansion or contraction of the number of repeated units that they contain, and this relatively high mutation rate has made them useful as markers in genetic linkage and forensic studies [7]. Repeat sequences that cause disease

when expanded beyond a threshold length are a specific instance of microsatellite instability. Many of these diseases are neurological, though why this should be so is unclear, although microsatellite instability is also observed in neoplasia [8–10]. HD is one of a group of neurological diseases caused by CAG repeat expansion above a threshold length. Repeat expansion can occur in germline and non-germline (somatic) cells with different consequences. Intergenerational repeat expansion through the germline underpins genetic anticipation, where disease onset occurs earlier in successive generations [11]. Even in diseases where it is hard to measure the length of the repeat disease, anticipation occurs, suggesting that germline expansion of the repeat also occurs [12]. Sperm analysis of repeat lengths in HD shows wide variation with a trend towards expansion rather than contraction, many mildly expanded alleles and a few very long expansions [13–16]. This is also seen in multiple mouse lines where expansion occurs over many generations leading to much longer alleles than those originally in the founders [17]. In HD, as in most other CAG repeat diseases and in microsatellite transmission in general, germline expansion is most marked on paternal transmission of the mutated allele, for unclear reasons [13, 18, 19].

Somatic expansion of repeats occurs in many different cell types, including both those that divide and those that are terminally differentiated (such as neurons), and is observed in many repeat expansion diseases [11, 20, 21]. In HD, mouse models and human post-mortem brain analyses have shown that somatic CAG repeat expansion occurs over time in a tissue-dependent manner, and is inversely correlated with age at disease onset [22]. The marked somatic expansion observed in striatum has led to the hypothesis that repeat expansion in striatal medium spiny neurons drives HD pathogenesis. However, large expansions also occur in some non-neural tissues such as liver which are not obviously involved in HD [21, 23]. Disease-associated somatic instability is modulated by the DNA damage response, and particularly by the mismatch repair system [9, 24, 25].

Therefore, by assaying repeat length and stability we can gain insight into the role of DNA repair in HD. Here we detail methods for CAG repeat sizing, and the cellular and mouse models that have been used to explore DNA repair in HD.

1.3. Methods for measuring CAG repeat length

Assessing the sequence and organisation of DNA in repetitive genomic regions is difficult. Various methods have been developed over the last 25 years to probe the CAG repeat in *HTT*, and these are outlined in Table 1. The most straightforward methods utilise PCR amplification of the repeat but there are inherent errors such as amplification bias towards smaller alleles, and ‘PCR stutter’ arising from DNA synthesis in a repetitive region. Non-amplification methods such as genomic DNA

digestion and Southern blotting avoid PCR bias but are laborious, require much more input DNA, and are semi-quantitative at best. In addition, most methods assess repeats in bulk reactions, i.e., utilising input DNA from thousands of different cells (from tissue or culture) at once. Large changes in repeat number can be observed, but rare alleles in individual cells will be missed and calculating repeat length averages is not informative for subtle effects. Input DNA dilution as in small-pool PCR can obviate these issues but is labour-intensive and prone to contamination. Next-generation sequencing (NGS) technologies using typical short-reads (up to 100 bp) can be used to assess wild-type alleles with few CAG repeats, but alignment of alleles containing expanded repeats is usually difficult, if not impossible. New long-read NGS offers exciting opportunities to revolutionise repeat sequence and length assessment: by generating sequencing reads of up to 80 kbp long repeats can be sequenced, sized and phased in one reaction, and there is the potential to multiplex many samples to increase throughput[26]. Currently most of these NGS methods work on amplicons across repeats, but techniques where repeats can be sequenced without amplification are in development. Costs are high, but are likely to decrease rapidly as the new technology becomes more widespread and embedded into standard analyses.

[Insert Table 1 about here]

2. Cellular models to assess CAG repeat stability *in vivo*

All methods for accurately sizing CAG repeats require DNA extraction from cells or tissue and subsequent analysis (Table 1). Although these assays can give an estimate of repeat length at a specific time point, they do not provide any information about the intracellular dynamics of the repeat instability reactions. To assess these repeat dynamics various models have been developed in bacterial, yeast, mammalian, and cell-free systems to try to provide insight into CAG repeat instability (Table 2). Most of these models utilise the fact that longer CAG repeats can interfere with gene expression and/or splicing as the basis for a selectable reporter system. Frameshifting assays are not useful here as the loss or gain of trinucleotide units leaves the translational frame unaffected. The reporter system can be plasmid-based or, more usefully, integrated into the host cell genome. Most recently a GFP minigene reporter system has been developed with the ability to identify expansions and contractions in the same population of living cells – this could provide an amenable system for testing DNA repair variants isolated in genetic studies [27]. All of the assay systems described in Table 2 highlight the inherent instability of the CAG repeat, the length threshold of 30-40 repeats seen in many human diseases, and show that contractions and expansions can occur in many genomic contexts. The systems are much more sensitive than PCR

methods for identifying rare repeat instability events (the cellular reporter systems act as biosensors to enrich for contractions/expansions) but often require PCR validation of results. DNA repair gene knockouts and variants have been assessed effectively in these systems [27, 28]. However, interpretation of results is limited by the uncertain relevance of the cell lines to human disease: cells may be non-human, they are usually dividing in culture, often immortalised, and the CAG repeat arrays are usually within an artificial reporter construct rather than a physiologically relevant context. Many long CAG repeats show a bias towards contraction in dividing cellular models in contrast to HD neurons *in vivo* where expansions are favoured. In addition, the time-frame of cellular experiments is usually days-weeks which is insignificant in the context of the human HD time-course that is measured in years.

[Insert Table 2 about here]

Primary and embryonic stem cells from mouse models of HD and human induced pluripotent stem cells from HD patients have also been cultured extensively and repeat lengths assessed[29–31]. These cells have the advantage of disease-relevant repeat contexts but require downstream *in vitro* repeat length analysis and are harder to manipulate genetically than the reporter systems above. In the future a reporter system that can be deployed in a multi-well plate design with automated readouts to show repeat instability in living cells, and ideally in cells derived from HD patients, would be the ultimate goal.

3. HD mouse models and DNA repair

Many transgenic and knock-in mouse models of HD have been generated since the discovery of the causative CAG expansion in *HTT* (reviewed in chapter by Bates [32]. These models recapitulate some of the characteristics of human disease and have been used to investigate the role of DNA repair processes in repeat stability and correlated phenotypes. Although very useful for assessing the roles of genes and mutations at the level of the whole organism, these models all have limitations in their representation of human HD. For example, mice have short lifespans and do not develop a disease phenotype unless genetically modified; most of the mice used have very long CAG repeats (>100) in order to drive a measurable phenotype in experimental time-frames whereas >90% of human HD patients have 40-50 CAG repeats. In addition, although many DNA repair genes are well conserved from mice to humans the increased complexity of the DNA damage response in humans, and the need to respond to DNA damage accumulated in neurons over decades (compared with months in mice), implies that results from mice will not necessarily hold in humans. However, multiple lines of

evidence have implicated DNA repair in CAG repeat instability and HD pathogenesis and so many repair genes have been tested in mouse models.

Length and age-dependent somatic CAG repeat expansion in striatal neurons of murine HD models correlates with worsening motor and behavioural phenotypes in the animals. However, it is extremely difficult to distinguish between pathology arising from the starting (long) repeat and that which may be related to somatic expansion of that repeat. Many studies have measured repeat stability in the context of DNA repair mutations to implicate DNA repair processes in HD (Table 3). There is most evidence for the involvement of the mismatch repair and base excision repair pathways: knockout of various genes in various HD models prevents CAG repeat expansion in both germline and striatum, and may be correlated with improved phenotype. There is some specificity of factors within these pathways: for example, *Msh2* and *Msh3* knockouts differ in somatic and germline effects; *Ogg1* and *Neil1* affect CAG repeats whereas other glycosylases (e.g. *Mpg*) do not. These findings suggest that there may be specific effects of certain repair enzymes on CAG repeats. One further difficulty with data interpretation arises as DNA repair mutant mice may have other deleterious phenotypes. For example, *Msh2* null mice show methylation tolerance and predisposition to lymphomas [33]. Nevertheless, the weight of evidence suggests that DNA repair enzymes do affect CAG repeats in mouse models of HD, and may be involved in pathogenesis.

[Insert Table 3 about here]

4. Discussion and future directions

There is now considerable evidence linking DNA repair with HD. Repair pathways are most likely to intersect with HD at the level of the CAG repeat, but mechanistic detail remains elusive. For example, it is unclear whether DNA repair can stimulate repeat expansion in an unbroken DNA duplex, or whether pre-existing DNA damage (e.g. a single or double strand break) is required. DNA transactions that melt the duplex, such as transcription or replication, have been linked to repeat instability in specific systems but their relevance to human disease is uncertain [34]. How might DNA transcription or replication trigger repeat instability? Duplex unwinding exposes single-stranded DNA which is then susceptible to damage by exogenous or endogenous reagents leading to a requirement for DNA repair. Repair could then stimulate repeat instability. However, such DNA transactions also require an open chromatin conformation and recruitment of transcription/replication factors, and will cause local changes in DNA supercoiling, all of which may also affect the stability of CAG repeats. Experimental approaches to disentangle these and other possibilities involve inducing DNA damage,

measuring global effects in cells, and then, in the context of HD, measuring repeat stability. DNA damage can be induced with exogenous agents both non-specific (e.g., hydrogen peroxide to cause oxidative damage; topoisomerase I inhibitors such as camptothecin to induce single-strand breaks) and specific (e.g., CRISPR/Cas9 or Zinc-finger nucleases to induce targeted single or double strand breaks) to probe different aspects of DNA repair. DNA damage in cellular systems can be measured using comet or alkaline comet assays (for strand breaks), or γH2AX immunofluorescence (for double-strand breaks in non-dividing cells), and repair measured through unscheduled DNA synthesis (UDS[35]) – but none of these assays is particularly specific or quantitative[36]. In addition, any assays involving perturbation of DNA repair systems are prone to significant confounders such as neoplasia and genomic instability – a particular issue in cells (e.g. induced pluripotent stem cells) that may be prone to accumulation of genetic rearrangements in the laboratory.

Over 450 human genes are now implicated in the DNA damage response (DDR) and although there are linear pathways within the DDR – such as mismatch repair or base excision repair – it is becoming increasingly clear that there is considerable functional redundancy between factors in different pathways, and the interplay between them is complex[36]. Such safeguards are crucial for maintaining genomic stability *in vivo* but make experimental investigation much more difficult. In HD only a few DDR genes have been investigated so far. These were mainly chosen on theoretical grounds linking microsatellite instability and mismatch repair. Results have shown that these genes *can* influence CAG stability in HD models, but do not show whether they are relevant to human disease. Recent human genetic data are starting to link actual genetic modifiers of HD onset and progression to variants in DNA repair genes[1, 37]. Researchers can now introduce disease-modifiers identified from the actual HD population into their model systems of DNA repair and should be able to draw much stronger, disease-relevant, mechanistic conclusions. Developments of such assay systems will allow more efficient testing of potential therapeutic compounds that affect these mechanisms.

Tables

Table 1: *In vitro* methods for measuring CAG repeat length in Huntington's disease.

Method	Description	Advantages	Disadvantages	Refs
Fluorescence PCR	PCR amplification across CAG repeat using fluorescently tagged primers. Products separated by capillary gel electrophoresis and sized using GeneScan software	- High throughput - Accurate sizing (+/- 1 CAG) - Semi-quantitative	- Bulk method - Low sensitivity for detecting rare (<10%) alleles	[38, 39]
Triplet-primed PCR	PCR amplification of repeats using fluorescent forward primer and CAG-binding reverse primer to yield ladder of products. Products separated by capillary gel electrophoresis and sized using GeneScan software	- Good for assessing long repeats - Useful to determine true homozygosity	- Bulk method - Not quantitative	[40, 41]
Small-pool/single-molecule PCR	Serial dilution is used to limit input DNA to 0.5-200 genome equivalents in multiple independent PCRs across the CAG repeat. Products are typically separated by electrophoresis, Southern blotted and probed for the repeat.	- Detection of rare repeat alleles (particularly expansions) that would be missed in bulk PCR methods	- DNA contamination - Labour-intensive	[42, 43]
Southern Blotting	Genomic DNA is digested by restriction enzymes, blotted and probed with a labelled DNA fragment that hybridises specifically to the repeat-containing region	- No amplification artefacts - Useful to determine true homozygosity - Sensitive for long repeats that may not amplify in PCR	- Bulk method - Laborious - Requires large amounts of DNA (5-20 μg)	[44]
Sanger sequencing	PCR amplification across CAG repeat, separation of alleles by electrophoresis, DNA purification and sequencing	- Generates exact repeat sequence and length data for both alleles - Corroboration of other methods	- Cannot easily distinguish homozygous or similar alleles - Labour-intensive - Sequencing through repeats can be problematic	
Long-read next-generation sequencing (NGS)	Short-read NGS can sometimes be used to assess wild-type alleles, but sequence alignment is impossible for longer repeats. Long-read NGS approaches use much larger read lengths which enable high-throughput sequencing through long pathogenic repeats	- High-throughput and can be multiplexed - Generates exact repeat sequence and length data for both alleles - New technology will enable sequencing without amplification	- Very expensive - High error-rate (up to 15% currently) but mitigated by high coverage - Protocols not established	[26, 45, 46]

Table 2. Cell-based models of CAG repeat instability.

Model system	Advantages	Disadvantages	Refs
Bacteria (mainly *E. coli*). CAG:CTG arrays of various lengths either on a reporter plasmid or integrated into the bacterial genome (e.g. at the *lacZ* locus). Antibiotic resistance and colony PCR can be used to assess repeat stability in a variety of genetic backgrounds.	- Simple organism - Genetically tractable - DNA repair well characterised - Rapid data generation	- Few bacteria usually contain CAG repeats - Dividing bacterial cells - Relevance to human cells?	[47, 48]
Yeast (*S. cerevisiae*). CAG repeat arrays inserted between TATA box and ATG of *URA3* reporter gene and construct integrated into *LYS2* genomic locus. Repeat expansion reduces transcription of *URA3* which can be selected. Similar design enables repeat contraction to be detected.	- Simple organism and screening - Genetically tractable - Integrated in genome - Can measure expansion and contraction	- Dividing yeast cells - Only useful for small (<50) repeat ranges - Requires validation by repeat sizing	[49–51]
Yeast (*S. cerevisiae*). CAG repeats inserted in the *GAL1* promoter (between the UAS and the TATA box) of the forward selection marker *CAN1*. Construct integrated in genome. Large repeat expansions (>20) prevent transactivation of *CAN1* which is selected by growth on canavanine.	- Simple organism and screening - Can detect large expansions - Can quantify by growth characteristics.	- Dividing yeast cells - Only detects large expansions - Requires validation by repeat sizing	[28]
Mammalian (yeast shuttle vector reporter). Plasmid with a mammalian origin of replication and a yeast reporter gene system containing a CAG array in the promoter. Mammalian cells transfected and propagated, plasmid recovered and then transfected into yeast for selection (see above).	- Mammalian cells - Versatile - Can adapt for contractions or expansions	- Laborious - Plasmid-based - Requires passage through yeast for selection - Requires PCR confirmation	[52–54]
Mammalian (selectable reporter gene). CAG repeats cloned into intron of the *APRT* gene or the *HPRT* minigene. Transcription through long repeats (>32) leads to disrupted mRNA splicing and non-functional gene product which can be selected. The assay can been adapted to show contractions or expansions.	- Mammalian cells - Sensitive (10^{-6} vs 10^{-3} for PCR methods) - Threshold in assay fits biological data for repeat diseases	- Time-consuming - Plasmid-based (although can be integrated) - Immortalised, cycling cells - Artificial repeat context - Can only detect small changes in repeat number	[55–58]
Mammalian (fluorescent reporter gene). Inducible GFP minigene with intronic CAG repeats integrated into genome of HEK293 cells. Longer repeats inhibit correct splicing of GFP. Fluorescence inversely proportional to repeat length so can assess repeats in living cells. Has been adapted for repeat contractions and expansions.	- Semi-quantitative readout - Living cells - Sensitive (10^{-8}) - Rapid (days) - Can increase throughput with flow cytometry	- Immortalised, cycling cells - Artificial repeat context - Confounders that stimulate/repress GFP production - Cannot detect very small changes in repeat number	[27, 59]
Mammalian (cell-free). Yeast shuttle vector reporter plasmid with $(CAG)_{22}$ array in promoter incubated with human cell-free extracts (HeLa) and then transfected into yeast. Expansion of ≥4 repeats leads to impaired *CAN1* expression and consequent canavanine resistance.	- Simple system - Contractions and expansions can be assayed - No mammalian transcription - Yeast used as biosensor to increase sensitivity	- Laborious - Plasmid-based - Requires passage through yeast for selection - Requires PCR confirmation	[60]

Table 3. Mouse models used to assess the role of DNA repair in Huntington's disease.

DNA repair pathway	DNA repair gene (murine)	HD mouse model(s)	Insights into DNA repair and HD from model system	Refs
Mismatch Repair	*Msh2*	R6 transgenics Q111 knock-in	Homozygous *Msh2* knockout prevents somatic and germline CAG repeat expansion in all models tested. Heterozygotes are unaffected. Conditional striatal knockout correlates absent repeat expansion with delayed neuropathology.	[61–64]
	Msh3	R6/1 transgenic Q111 knock-in	Homozygous *Msh3* knockout prevents somatic but not germline CAG repeat expansion in Q111 mice. Heterozygotes also show much reduced repeat expansion. Decreased MSH3 protein expression/stability correlates with repeat stability in R6/1 mice.	[43, 65]
	Msh6	Q111 knock-in	Heterozygous *Msh6* knockout increases CAG repeat contractions on paternal transmission, but homozygous knockout is as wild-type. Inconsistent data. Somatic instability of repeats unaffected.	[65]
	Mlh1	Q111 knock-in	Homozygous *Mlh1* knockout prevents somatic CAG repeat expansion. Heterozygotes are unaffected. *Mlh1* expression levels correlate with repeat instability in different mouse backgrounds.	[66]
	Mlh3	Q111 knock-in	Homozygous *Mlh3* knockout prevents somatic CAG repeat expansion. Heterozygotes show reduction in repeat expansion.	[66]
Base excision repair	*Ogg1*	R6/1 transgenic Q150 knock-in	Homozygous *Ogg1* knockout prevents age-dependent somatic CAG repeat expansion in ~70% of R6/1 HD mice. Loss of somatic expansion by knockout of *Ogg1* in Q150 HD mice correlates with significant delay in onset of motor decline (rotarod).	[67, 68]
	Neil1	R6/1 transgenic	Homozygous *Neil1* knockout associated with reduction of germline and somatic CAG instability. Not tissue specific. Small numbers.	[69]
	Mpg	R6/1 transgenic	Homozygous *Mpg* knockout has no effect on somatic stability of CAG repeat.	[67]
	Nthl1	R6/1 transgenic	Homozygous *Nthl1* knockout has no effect on somatic stability of CAG repeat.	[67]
	PolB	R6 transgenics	POLβ protein specifically enriched at CAG repeats in striatum (detected by chromatin immunoprecipitation) and increasing enrichment with age.	[70]
Nucleotide excision repair	*Csb*	R6/1 transgenic	Homozygous *Csb* knockout increases germline CAG expansions but has little effect on somatic CAG repeats. *Csb* -/- *Ogg1* -/- double mutant shows increased somatic CAG expansion. Small numbers and limited data only.	[71]
	Xpc	Q111 knock-in	Homozygous *Xpc* knockout has no effect on germline or striatal CAG repeat stability	[65]
Single-strand break repair	*Parp-1*	R6/2 transgenic	Inhibition of PARP-1 by intraperitoneal injection of INO-1001 improves motor coordination (rotarod) and survival of R6/2 mice. Improvements correlate with less striatal neuropathology. Unclear how effects are mediated. CAG repeats not examined.	[72]
Double-strand break repair	*Ku70*	R6/2 transgenic	Double-strand DNA breaks accumulate in striatal neurons of R6/2 mice. KU70 binds mutant HTT protein. Overexpression of KU70 improves clasping function and survival of R6/2 mice. Small numbers. CAG repeats not examined.	[73]
	Atm	BACHD transgenic	Increased ATM signalling in HD mouse striatum (and in human brain tissue). Heterozygous *Atm* null BACHD mice show improved motor abilities and less neuropathology (mirrored by ATM inhibition by KU-60019 in HD cell model).	[74]

References

1. Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium (2015) Identification of genetic factors that modify clinical onset of Huntington's disease. Cell 162:516–526
2. Moss DJH, Pardiñas AF, Langbehn D et al (2017) Identification of genetic variants associated with Huntington's disease progression: a genome-wide association study. Lancet Neurol 16:701–711
3. Stone, Timothy C; Holmans PA (2018) Using genomic data to find disease-modifying loci in HD. In Methods in Mol Biol.: Huntington's Disease; eds Precious, SV, Rosser, AE & Dunnett, SB. Humana Press, Totowa NJ, USA. In press
4. Bettencourt C, Hensman-Moss D, Flower M et al (2016) DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases. Ann Neurol 79:983–990
5. Jones L, Houlden H, Tabrizi SJ (2017) DNA repair in the trinucleotide repeat disorders. Lancet Neurol 16:88–96
6. Ashkenazi A, Bento CF, Ricketts T et al (2017) Polyglutamine tracts regulate beclin 1-dependent autophagy. Nature 545:108–111
7. Edwards A, Caskey CT (1991) Genetic marker technology. Curr Opin Biotechnol 2:818–822
8. Trang H, Stanley SY, Thorner P et al (2015) Massive CAG repeat expansion and somatic instability in maternally transmitted infantile spinocerebellar ataxia type 7. JAMA Neurol 72:219–23
9. Schmidt MHM, Pearson CE (2016) Disease-associated repeat instability and mismatch repair. DNA Repair (Amst) 38:117–126
10. Thibodeau SN, Bren G, Schaid D (1993) Microsatellite instability in cancer of the proximal colon. Science 260:816–819
11. Nelson DL, Orr HT, Warren ST (2013) The unstable repeats - three evolving faces of neurological disease. Neuron 77:825–43
12. Van Mossevelde S, van der Zee J, Gijselinck I et al (2017) Clinical evidence of disease anticipation in families segregating a C9orf72 repeat expansion. JAMA Neurol 74:445-452
13. Telenius H, Kremer B, Goldberg YP et al (1994) Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm. Nat Genet 6:409–414
14. Telenius H, Almqvist E, Kremer B et al (1995) Somatic mosaicism in sperm is associated with intergenerational (CAG)n changes in Huntington disease. Hum Mol Genet 4:189–195
15. Leeflang EP, Zhang L, Tavare S et al (1995) Single sperm analysis of the trinucleotide repeats in the Huntington's disease gene: quantification of the mutation frequency spectrum. Hum Mol Genet 4:1519–1526
16. Semaka A, Kay C, Doty C et al (2013) CAG size-specific risk estimates for intermediate allele repeat instability in Huntington disease. J Med Genet 50:696–703
17. Larson E, Fyfe I, Morton AJ et al (2015) Age-, tissue- and length-dependent bidirectional somatic CAG*CTG repeat instability in an allelic series of R6/2 Huntington disease mice. Neurobiol Dis 76:98–111
18. Kovtun I V, Welch G, Guthrie HD et al (2004) CAG repeat lengths in X- and Y-bearing sperm indicate that gender bias during transmission of Huntington's disease gene is determined in the embryo. J Biol Chem 279:9389–9391
19. Brinkmann B, Klintschar M, Neuhuber F et al (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. Am J Hum Genet 62:1408–1415
20. Orr HT, Zoghbi HY (2007) Trinucleotide repeat disorders. Annu Rev Neurosci 30:575–621
21. Kennedy L, Evans E, Chen C-M et al (2003) Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. Hum Mol Genet 12:3359–3367

22. Swami M, Hendricks AE, Gillis T et al (2009) Somatic expansion of the Huntington's disease CAG repeat in the brain is associated with an earlier age of disease onset. Hum Mol Genet 18:3039–47

23. Shelbourne PF, Keller-McGandy C, Bi WL et al (2007) Triplet repeat mutation length gains correlate with cell-type specific vulnerability in Huntington disease brain. Hum Mol Genet 16:1133–1142

24. Iyer RR, Pluciennik A, Napierala M et al (2015) DNA triplet repeat expansion and mismatch repair. Annu. Rev Biochem 84:199–226

25. Zhao X-N, Usdin K (2015) The Repeat Expansion Diseases: the dark side of DNA repair? DNA Repair (Amst) 32:96–105

26. Rhoads A, Au KF (2015) PacBio sequencing and its applications. Genomics. Proteomics Bioinformatics 13:278–89

27. Cinesi C, Aeschbach L, Yang B et al (2016) Contracting CAG/CTG repeats using the CRISPR-Cas9 nickase. Nat Commun 7:13272

28. Kim JC, Harris ST, Dinter T et al (2016) The role of break-induced replication in large-scale expansions of (CAG)n/(CTG)n repeats. Nat Struct Mol Biol 24:55-60

29. Jonson I, Ougland R, Klungland A et al (2013) Oxidative stress causes DNA triplet expansion in Huntington's disease mouse embryonic stem cells. Stem Cell Res 11:1264–71

30. Gomes-Pereira M, Monckton DG (2004) Mouse tissue culture models of unstable triplet repeats. Methods Mol Biol 277:215–227

31. Seriola A, Spits C, Simard JP et al (2011) Huntington's and myotonic dystrophy hESCs: down-regulated trinucleotide repeat instability and mismatch repair machinery expression upon differentiation. Hum Mol Genet 20:176–185

32. Bates GP (2018) HD mice: genetic and knock-in models. In Methods in Mol Biol.: Huntington's Disease; eds Precious, SV, Rosser, AE & Dunnett, SB. Humana Press, Totowa NJ, USA. In press

33. de Wind N, Dekker M, Berns A et al (1995) Inactivation of the mouse Msh2 gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination, and predisposition to cancer. Cell 82:321–30

34. Lin Y, Wilson JH (2012) Nucleotide excision repair, mismatch repair, and R-loops modulate convergent transcription-induced cell death and repeat instability. PLoS One 7:e46807

35. Sepe S, Milanese C, Gabriels S et al (2016) Inefficient DNA repair Is an aging-related modifier of Parkinson's disease. Cell Rep 15:1866–1875

36. Pearl LH, Schierz AC, Ward SE et al (2015) Therapeutic opportunities within the DNA damage response. Nat Rev Cancer 15:166–80

37. Moss DJH, Pardinas AF, Langbehn D et al (2017) Identification of genetic variants associated with Huntington's disease progression: a genome-wide association study. Lancet Neurol 16:701-711

38. Vnencak-Jones CL (2003) Fluorescence PCR and GeneScan analysis for the detection of CAG repeat expansions associated with Huntington's disease. Methods Mol Biol 217:101–8

39. Lee J-M, Zhang J, Su AI et al (2010) A novel approach to investigate tissue-specific trinucleotide repeat instability. BMC Syst Biol 4:29

40. Warner JP, Barron LH, Goudie D et al (1996) A general method for the detection of large CAG repeat expansions by fluorescent PCR. J Med Genet 33:1022–6

41. Jama M, Millson A, Miller CE et al (2013) Triplet repeat primed PCR simplifies testing for Huntington disease. J Mol Diagnostics 15:255–262

42. Gomes-Pereira M, Bidichandani SI, Monckton DG (2004) Analysis of unstable triplet repeats using small-pool polymerase chain reaction. Methods Mol Biol 277:61–76

43. Tomé S, Manley K, Simard JP et al (2013) MSH3 polymorphisms and protein levels affect CAG repeat instability in Huntington's disease mice. PLoS Genet 9:e1003280

44. Guida M, Fenwick RG, Papp AC et al (1996) Southern transfer protocol for confirmation of Huntington disease. Clin Chem 42:1711–2

45. Loomis EW, Eid JS, Peluso P et al (2013) Sequencing the unsequenceable: expanded CGG-repeat alleles of the fragile X gene. Genome Res 23:121–8

46. Doi K, Monjo T, Hoang PH et al (2014) Rapid detection of expanded short tandem repeats in personal genomics using hybrid sequencing. Bioinformatics 30:815–22

47. Bichara M, Wagner J, Lambert IB (2006) Mechanisms of tandem repeat instability in bacteria. Mutat. Res. Mol Mech Mutagen 598:144–163

48. Zahra R, Blackwood JK, Sales J et al (2007) Proofreading and secondary structure processing determine the orientation dependence of CAG x CTG trinucleotide repeat instability in Escherichia coli. Genetics 176:27–41

49. Dixon MJ, Bhattacharyya S, Lahue RS (2004) Genetic assays for triplet repeat instability in yeast. In: Trinucleotide Repeat Protoc. Humana Press, New Jersey, pp 029–046

50. Miret JJ, Pessoa-Brandao L, Lahue RS (1998) Orientation-dependent and sequence-specific expansions of CTG/CAG trinucleotide repeats in Saccharomyces cerevisiae. Proc Natl Acad Sci USA 95:12438–12443

51. Williams GM, Surtees JA (2015) MSH3 promotes dynamic behavior of trinucleotide repeat tracts in vivo. Genetics 200:737–754

52. Pelletier R, Farrell BT, Miret JJ et al (2005) Mechanistic features of CAG•CTG repeat contractions in cultured cells revealed by a novel genetic assay. Nucleic Acids Res 33:5667

53. Farrell BT, Lahue RS (2006) CAG*CTG repeat instability in cultured human astrocytes. Nucleic Acids Res 34:4495–505

54. Claassen DA, Lahue RS (2007) Expansions of CAG.CTG repeats in immortalized human astrocytes. Hum Mol Genet 16:3088–96

55. Gorbunova V, Seluanov A, Dion V et al (2003) Selectable system for monitoring the instability of CTG/CAG triplet repeats in mammalian cells. Mol Cell Biol 23:4485–4493

56. Lin Y, Dion V, Wilson JH (2005) A novel selectable system for detecting expansion of CAG.CTG repeats in mammalian cells. Mutat Res 572:123–131

57. Lin Y, Dion V, Wilson JH (2006) Transcription promotes contraction of CAG repeat tracts in human cells. Nat Struct. Mol Biol 13:179–180

58. Lin Y, Wilson JH (2007) Transcription-induced CAG repeat contraction in human cells is mediated in part by transcription-coupled nucleotide excision repair. Mol Cell Biol 27:6209–6217

59. Santillan BA, Moye C, Mittelman D et al (2014) GFP-based fluorescence assay for CAG repeat instability in cultured human cells. PLoS One 9:e113952

60. Stevens JR, Lahue EE, Li G-M et al (2013) Trinucleotide repeat expansions catalyzed by human cell-free extracts. Cell Res 23:565–72

61. Manley K, Shirley TL, Flaherty L et al (1999) Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice. Nat Genet 23:471–3

62. Kovtun I V., McMurray CT (2001) Trinucleotide expansion in haploid germ cells by gap repair. Nat Genet 27:407–411

63. Kovalenko M, Dragileva E, St Claire J et al (2012) Msh2 acts in medium-spiny striatal neurons as an enhancer of CAG instability and mutant huntingtin phenotypes in Huntington's disease knock-in mice. PLoS One 7:e44273

64. Wheeler VC, Lebel L-A, Vrbanac V et al (2003) Mismatch repair gene Msh2 modifies the timing of early disease in HdhQ111 striatum. Hum Mol Genet 12:273–281

65. Dragileva E, Hendricks A, Teed A et al (2009) Intergenerational and striatal CAG repeat instability in Huntington's disease knock-in mice involve different DNA repair genes. Neurobiol Dis 33:37–47

66. Pinto RM, Dragileva E, Kirby A et al (2013) Mismatch repair genes Mlh1 and Mlh3 modify CAG instability in Huntington's disease mice: genome-wide and candidate approaches. PLoS Genet 9:e1003930

67. Kovtun I V, Liu Y, Bjoras M et al (2007) OGG1 initiates age-dependent CAG trinucleotide expansion in somatic cells. Nature 447:447–52

68. Budworth H, Harris FR, Williams P et al (2015) Suppression of somatic expansion delays the onset of pathophysiology in a mouse model of Huntington's disease. PLoS Genet 11:e1005267

69. Møllersen L, Rowe AD, Illuzzi JL et al (2012) Neil1 is a genetic modifier of somatic and germline CAG trinucleotide repeat instability in R6/1 mice. Hum Mol Genet 21:4939–47

70. Goula A-V, Berquist BR, Wilson DM et al (2009) Stoichiometry of base excision repair proteins correlates with increased somatic CAG instability in striatum over cerebellum in Huntington's disease transgenic mice. PLoS Genet 5:e1000749

71. Kovtun I V., Johnson KO, McMurray CT (2011) Cockayne syndrome B protein antagonizes OGG1 in modulating CAG repeat length in vivo. Aging (Albany. NY) 3:509–514

72. Cardinale A, Paldino E, Giampà C et al (2015) PARP-1 inhibition is neuroprotective in the R6/2 mouse model of Huntington's dDisease. PLoS One 10:e0134482

73. Enokido Y, Tamura T, Ito H et al (2010) Mutant huntingtin impairs Ku70-mediated DNA repair. J Cell Biol 189:425–43

74. Lu X-H, Mattis VB, Wang N et al (2014) Targeting ATM ameliorates mutant Huntingtin toxicity in cell and animal models of Huntington's disease. Sci Transl Med 6:268ra178