

## REVIEW

# The central role of DNA damage and repair in CAG repeat diseases

Thomas H. Massey and Lesley Jones\*

## ABSTRACT

Diseases such as Huntington's disease and certain spinocerebellar ataxias are caused by the expansion of genomic cytosine-adenine-guanine (CAG) trinucleotide repeats beyond a specific threshold. These diseases are all characterised by neurological symptoms and central neurodegeneration, but our understanding of how expanded repeats drive neuronal loss is incomplete. Recent human genetic evidence implicates DNA repair pathways, especially mismatch repair, in modifying the onset and progression of CAG repeat diseases. Repair pathways might operate directly on repeat sequences by licensing or inhibiting repeat expansion in neurons. Alternatively, or in addition, because many of the genes containing pathogenic CAG repeats encode proteins that themselves have roles in the DNA damage response, it is possible that repeat expansions impair specific DNA repair pathways. DNA damage could then accrue in neurons, leading to further expansion at repeat loci, thus setting up a vicious cycle of pathology. In this review, we consider DNA damage and repair pathways in postmitotic neurons in the context of disease-causing CAG repeats. Investigating and understanding these pathways, which are clearly relevant in promoting and ameliorating disease in humans, is a research priority, as they are known to modify disease and therefore constitute prevalidated drug targets.

**KEY WORDS:** CAG repeat, DNA damage, DNA repair, Huntington's disease, Spinocerebellar ataxia

## Introduction

Expanded cytosine-adenine-guanine (CAG) trinucleotide repeats in the exons of certain genes can induce neurodegeneration in the central nervous system (CNS). Diseases caused by expanded CAG repeats include Huntington's disease (HD), which has a prevalence of approximately 1 in 8000 in populations of European descent (Evans et al., 2013; Fisher and Hayden, 2014), and various spinocerebellar ataxias (SCAs), which are individually very rare but have a combined prevalence of ~1 in 40,000 in European/Asian populations (Ruano et al., 2014). These dominantly inherited diseases are all characterised by slow, progressive neuronal loss over 10–20 years, leading to worsening disability and, eventually, death. Specific clinical manifestations depend on the genes and cell types affected by the repeat expansion but, despite the phenotypic variation between these diseases, a common underlying molecular

pathology seems likely. In support of this hypothesis, recent human genetic data suggest that DNA repair pathways are central to the pathogenesis of CAG repeat diseases (see Glossary, Box 1) (Bettencourt et al., 2016; GeM-HD Consortium, 2015). In this review, we consider how DNA damage and repair pathways could potentially mediate CAG repeat-driven pathology in CNS neurons. A better understanding of these cellular mechanisms could identify novel drug targets, an urgent requirement in the field given that there are currently no disease-modifying treatments for any CAG repeat disorder.

## DNA damage and repair in the CNS

DNA is continually damaged and repaired in all living cells. Intricate DNA repair mechanisms have evolved in parallel with increasing genome complexity in order to preserve genetic information (O'Brien, 2006). However, inaccurate repair can be mutagenic, while failed repair can threaten the integrity of the genome. Different cells sustain different types of DNA damage, and various overlapping mechanisms within the overarching DNA damage response (DDR; see Glossary, Box 1) are required for effective repairs (Jackson and Bartek, 2009; Pearl et al., 2015). The adult postmitotic neurons that degenerate in CAG repeat diseases sustain and repair particular types of DNA damage, as we discuss below.

## DNA damage

It has been estimated that each mammalian cell sustains as many as 10,000 single-strand and 10–50 double-strand DNA breaks per day (Madabhushi et al., 2014). Exogenous sources of DNA damage, such as UV light, ionising radiation and chemical mutagens, predominantly affect exposed and dividing cells, but in neurons, and especially in neurons of the CNS, endogenous metabolic processes are more relevant sources of DNA damage. The high oxygen demands of the brain (20% of total body oxygen consumption, but only 2% of body mass) expose its cells to numerous reactive oxygen species (ROS; see Glossary, Box 1) produced by normal mitochondrial respiration (Cooke et al., 2003), and further oxidative damage can arise as a result of inflammation. Over 100 different types of DNA base damage have been identified as being caused by ROS, the most abundant of which is 8-oxo-2'-deoxyguanosine (8-oxo-dG). One study of human lymphocytes found 8-oxo-dG at a steady state of ~10,000 damaged bases per cell nucleus (Ohno et al., 2006). This altered base can be premutagenic in replicating cells and has inhibitory effects on transcription in postmitotic neurons (Iyama and Wilson, 2013). In addition to ROS, cellular metabolism generates endogenous alkylating compounds (e.g. S-adenosyl methionine), lipid peroxidation products, and reactive nitrogen and carbonyl species that can directly damage DNA. DNA is also susceptible to hydrolysis, which can directly cause base loss or deamination, particularly in single-stranded regions. CAG repeats might be especially susceptible to damage as

Institute of Psychological Medicine and Clinical Neurosciences, MRC Centre for Neuropsychiatric Genetics and Genomics, Hadyn Ellis Building, Cardiff University, Cardiff, CF24 4HQ, UK.

\*Author for correspondence (JonesL1@cf.ac.uk)

 T.H.M., 0000-0002-9804-2131; L.J., 0000-0002-3007-4612

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0/>), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

**Box 1. Glossary**

**Autophagy:** mechanism by which cells recycle macromolecules through lysosomes.

**Base excision repair (BER):** pathway that senses and repairs small, nondistorting base lesions in DNA (e.g. arising from oxidative damage). Two major subpathways are known: short-patch (SP-BER) and long-patch (LP-BER) relating to the amount of gap-filling DNA repair synthesis required (Fig. 2).

**CAG repeat disorder:** disease caused by a number of repeated, consecutive CAG trinucleotide units in DNA over a threshold length.

**Cockayne syndrome B protein (CSB):** ATPase with multiple functions in DNA repair as well as roles in chromatin remodelling, transcription and mitochondrial function. Mutations in CSB cause ~75% of Cockayne syndrome, a disease of neurodevelopmental abnormalities, neurodegeneration and premature ageing.

**DNA damage response (DDR):** network of overlapping pathways in cells involved with DNA damage signalling and repair, integrated with the cell cycle. Encoded by >450 genes in humans.

**Double-strand break repair (DSBR):** mechanism for ensuring genomic integrity following double-strand DNA breakage. Homologous recombination and nonhomologous end joining are the two main pathways (Fig. 1).

**Genetic modifier:** genetic variant that is not directly causative for a disease, but can affect the phenotype when occurring together with the disease-causing mutation.

**Genome-wide association study (GWAS):** unbiased, observational, pan-genome screen for common genetic variants associated with a particular disease or trait.

**Lynch syndrome:** a cancer predisposition syndrome caused by mutations in mismatch repair genes such as *MSH2* and *MLH1*. Mutation carriers are at increased risk of colorectal and other cancers.

**Medium spiny neurons (MSNs):** inhibitory GABA-ergic interneurons that make up >95% of striatal neurons in the human brain. First neurons to degenerate in Huntington's disease.

**Mismatch repair (MMR):** pathway that is canonically involved in strand-specific repair of mismatched base-pairs arising from DNA replication errors in proliferating cells. Recent data indicate broader repair functions, including in nondividing cells such as neurons (Fig. 2).

**MutL complex:** protein dimer involved in mismatch processing downstream of MutS complex in the MMR pathway (Fig. 2).

**MutS complex:** protein dimer responsible for initial mismatch recognition in MMR. Two complexes with overlapping substrate specificities exist in eukaryotes: MutS $\alpha$  (MSH2/MSH6) and MutS $\beta$  (MSH2/MSH3), the latter found predominantly in neurons (Fig. 2).

**Nucleotide excision repair (NER):** a DNA repair pathway that senses and repairs bulky lesions (e.g. photoproducts from UV irradiation) by removing and replacing damaged nucleotides. Two major subpathways are known: global genomic (GG-NER), involved in pan-genomic DNA repair, and transcription-coupled (TC-NER), involved in repair in actively transcribed genes. These subpathways differ in damage sensing, but share downstream repair processes (Fig. 2).

**Poly(ADP-ribose) polymerases (PARP):** family of nuclear enzymes that detect single-strand breaks in DNA and signal to downstream repair factors through ADP ribosylation of target proteins.

**Reactive oxygen species (ROS):** byproducts of oxidative cellular metabolism that can react with, and damage, DNA and other macromolecules. Examples include superoxide radicals and hydrogen peroxide.

**Repeat-associated non-ATG (RAN) translation:** noncanonical mRNA translation initiated by tandem repeats rather than the ATG codon and leading to toxic homopolymeric proteins in cells.

**Single-strand break repair (SSBR):** a pathway that senses and repairs breaks in one strand of the DNA double helix. Shares components with base excision repair (Fig. 2).

**SPO11:** tyrosine recombinase that initiates recombination in meiosis I by inducing programmed double-strand breaks in DNA.

**Synthetic lethality:** interaction between two genes where deficiency of either alone is tolerated, but simultaneous deficiency of both is lethal to the cell. Harnessed in screens for novel (cancer) therapeutics.

**Topoisomerase:** enzyme that regulates the supercoiling of DNA by cleavage and re-ligation reactions on one strand (Type I) or both strands (Type II) of the double helix.

**V(D)J recombination:** a mechanism of programmed DNA strand breakage and repair that occurs in maturing B cells and T cells to generate antibody and T cell receptor diversity, respectively.

hydrolytic depurination (i.e. loss of A or G) and cytosine deamination are frequent events (Lindahl, 1993).

In addition to pathological DNA damage, there is increasing recognition of the role of physiological, 'programmed' DNA strand breakage. For example, outside the CNS, the generation of antibody and T cell receptor diversity depend on programmed double-strand

breaks, and their subsequent repair, as part of V(D)J recombination (see Glossary, Box 1) (Slean et al., 2008), and meiotic crossing over is initiated by SPO11-induced double-strand breaks (see Glossary, Box 1) (Keeney et al., 2014). Topoisomerases, which can introduce temporary single- or double-strand breaks in DNA to regulate supercoiling (see Glossary, Box 1), are essential for the expression of

long genes and are particularly important in the brain, where expressed genes are longer than elsewhere in the body (Gabel et al., 2015; King et al., 2013; Zylka et al., 2015). Recently, it has also been shown that neuronal activity in cells and in animals can trigger topoisomerase-induced double-strand DNA breaks in the promoters of neuronal early-response genes (Ju et al., 2006; Madabhushi et al., 2015; Suberbielle et al., 2013). These DNA breaks might relax topological constraints and stimulate promoter activity. Topoisomerase mutations can lead to various neurodevelopmental and neurodegenerative conditions, highlighting the importance of torsional regulation of DNA in brain function (King et al., 2013; McKinnon, 2016).

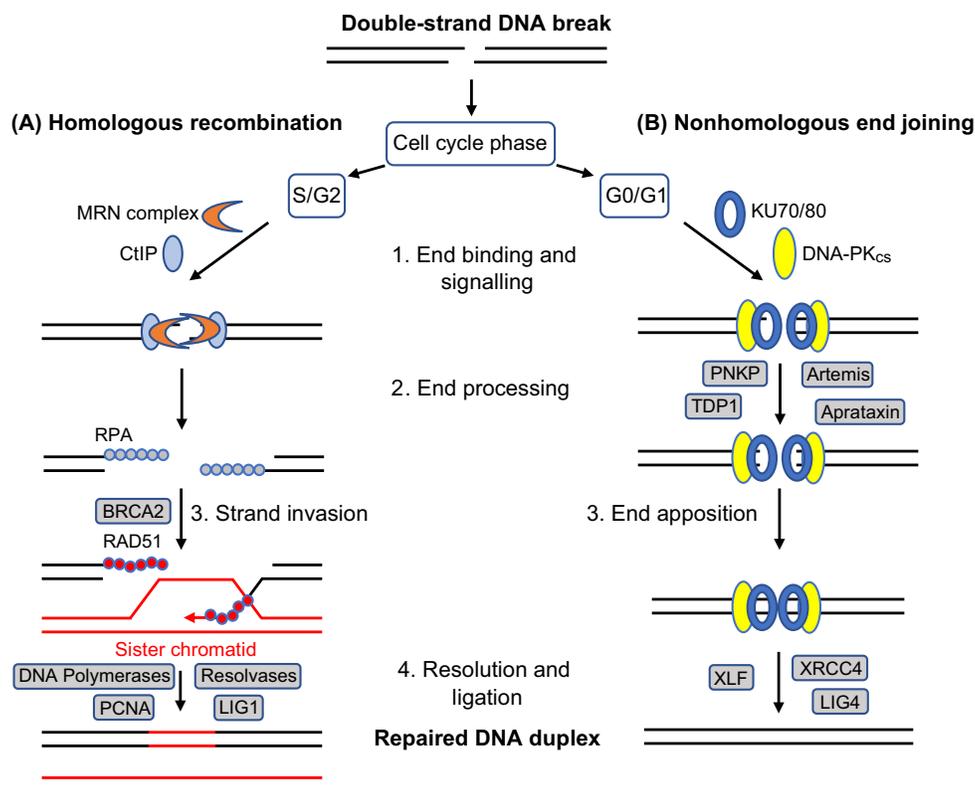
Therefore, neuronal DNA strands are continually broken and repaired *in vivo*. The tight regulation of DNA repair is crucial for maintaining neuronal gene expression and function; indeed, many Mendelian neurological diseases result directly from defects in the DNA damage response (Madabhushi et al., 2014).

**DNA repair**

Unrepaired DNA damage can have profound consequences on cells. For example, point mutations or chromosomal rearrangements can lead to cancer in dividing cells, such as glia, and can induce cell death in nondividing cells, such as CNS neurons. Moreover, lesions or strand breaks can stall DNA and RNA polymerases, leading to impaired replication or transcription, respectively, and potentially

triggering cell death or senescence (Hoeijmakers, 2009). Therefore, an elaborate damage response has evolved to identify and repair DNA damage in conjunction with cell cycle regulation. The DDR involves >450 genes in humans, with subsets of these genes deployed depending on the type of DNA damage, type of cell, stage of cell cycle and stage of organism development (Pearl et al., 2015).

Human CNS development involves the proliferation and subsequent migration and differentiation of neural progenitor cells, beginning a few weeks after conception and continuing to ~6 months after birth (Rulten and Caldecott, 2013). Most neurons then enter a postmitotic phase (and are required to survive for a lifetime), although there is clearly a high turnover of molecules within these cells. A small subset of neurons in the human hippocampus and lateral ventricle of the mature CNS can divide and contribute to ongoing neurogenesis (McKinnon, 2013). The phase of cell cycle determines which DNA repair pathways are utilised in neurons. For example, dividing progenitors in S/G2 phase use accurate homologous recombination (HR) for double-strand break repair (see Glossary, Box 1) and replication fork maintenance (Fig. 1). Mutations in these repair systems are embryonic lethal or lead to profound neurodevelopmental defects (McKinnon, 2013; Rulten and Caldecott, 2013). By contrast, mature postmitotic neurons in G0/G1 rely upon less accurate nonhomologous end joining (NHEJ) for double-strand break repair in the absence of a sister chromatid (Fig. 1). In addition, DNA damage in these neurons



**Fig. 1. DNA double-strand break repair pathways in neurons, highlighting key similarities and differences.** (A) Homologous recombination is utilised in S or G2 phase of dividing neuronal progenitors. DNA ends are processed by MRN complexes and other proteins to produce 3'-overhangs coated by RPA proteins. BRCA2 catalyses the exchange of RPA for RAD51, thus enabling invasion of the sister chromatid and error-free repair. (B) Nonhomologous end joining is utilised by postmitotic neurons in G0 or G1. DNA ends are bound by KU70/80, leading to the recruitment of DNA-PK<sub>cs</sub>. End processing is carried out by various enzymes including PNKP and Artemis (DCLRE1C), and then ends are ligated by LIG4-XRCC4. This form of repair preserves genomic integrity but can be error prone. Proteins at key commitment points are shown in colour, others in grey boxes. Some factors involved in double-strand break repair and cell-cycle regulation are omitted for clarity. BRCA2, breast cancer type 2 susceptibility protein; CtIP, C-terminal binding protein 1 interacting protein; DNA-PK<sub>cs</sub>, DNA-dependent protein kinase, catalytic subunit; MRN complex, MRE11-RAD50-NBS1 complex; LIG, DNA ligase; PCNA, proliferating cell nuclear antigen; PNKP, polynucleotide kinase 3'-phosphatase; RPA, replication protein A; TDP1, tyrosyl-DNA phosphodiesterase 1; XLF, XRCC4-like factor; XRCC4, X-ray repair cross-complementing 4.

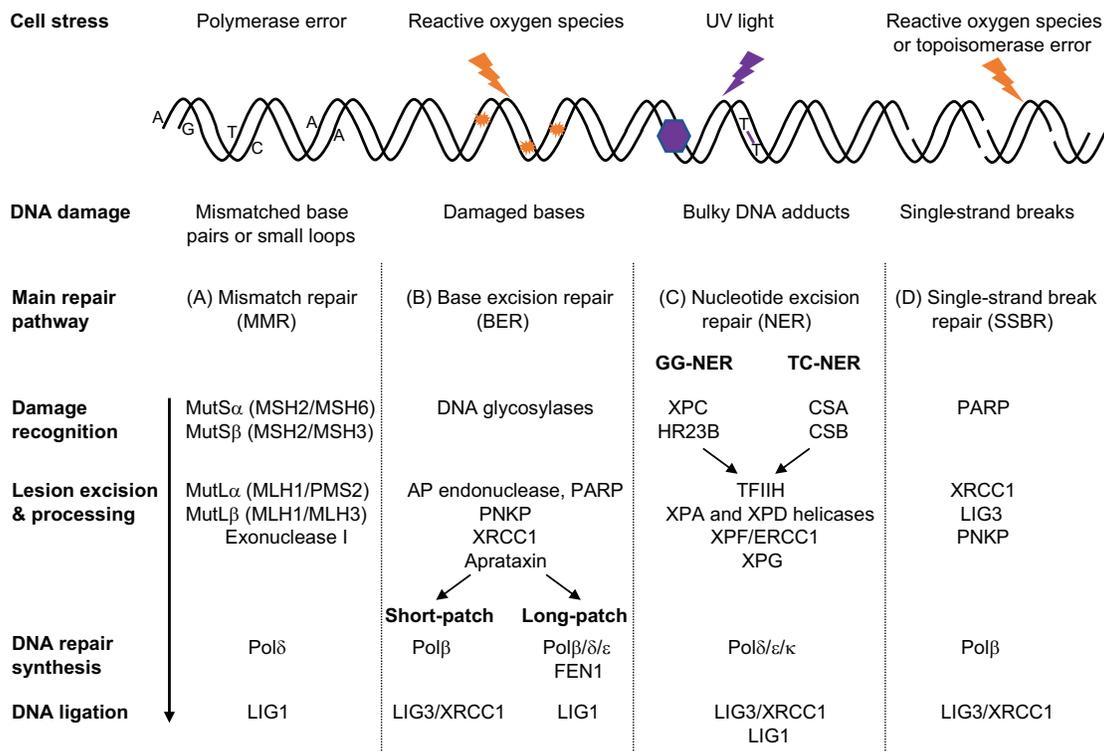
is predominantly single stranded, which can lead to the inhibition of transcription, as well as compromised genomic integrity. Pathways such as single-strand break repair (SSBR; see Glossary, Box 1) and transcription-coupled nucleotide excision repair (TC-NER; see Glossary, Box 1) are essential to cellular function, and mutations in these pathways lead to late neurodevelopmental defects, such as microcephaly, and, more commonly, neurodegeneration (McKinnon, 2013; Rulten and Caldecott, 2013). The type and location of affected neurons determine the clinical phenotypes observed (Madabhushi et al., 2014).

Many parts of the DDR are highly conserved from prokaryotes to eukaryotes, with extra layers of regulation and redundancy found in higher organisms. Individual linear repair pathways, such as base excision repair (BER; see Glossary, Box 1), nucleotide excision repair (NER; see Glossary, Box 1) and mismatch repair (MMR; see Glossary, Box 1), involve a series of analogous steps: lesion recognition, repair factor recruitment, lesion excision leading to DNA strand breakage, processing of DNA ends and DNA synthesis, to complete repair (Fig. 2). In addition, it is increasingly recognised that there is significant redundancy between pathways, presumably arising through divergent evolution in order to repair a wide range of lesions whilst maintaining genomic integrity (O'Brien, 2006; Pearl et al., 2015). Pathways can also be involved in noncanonical repairs. For example, MMR enzymes, which usually act at replication forks in dividing cells to correct DNA polymerase errors, have been

shown to recognise and repair mispaired bases in nondividing yeast cells (Rodriguez et al., 2012). Therefore, DNA repair systems have activities that depend not only on the type of DNA damage but also on the cell/tissue context, the background genetics of the cell, and the prevailing environmental conditions. CAG repeats in the DNA of adult postmitotic neurons constitute a particular substrate for repair systems, and the regulation of these repair activities is entwined with CAG repeat disease pathogenesis, as discussed below.

**Repetitive DNA and CAG repeat disorders**

Recent estimates suggest that >65% of the human genome consists of repetitive elements, ranging from microsatellites (2-6 base pair tandem repeats) up to whole genes (e.g. rDNA gene arrays) (Hall et al., 2017; de Koning et al., 2011). These elements can be coding or noncoding and have a range of essential functions in cells, including at centromeres and telomeres. Microsatellites are known to be common and hypermutable in both prokaryotic and eukaryotic genomes (Bichara et al., 2006). Their mutability can aid adaptation to changing environments, particularly in microorganisms, and might have a role in the regulation of gene expression in eukaryotes (Bichara et al., 2006; Gymrek et al., 2015). The processive mechanisms of DNA and RNA polymerases on unwound DNA mean that tandem repeats can readily adopt noncanonical conformations in DNA, such as slipped strands, hairpin loops,



**Fig. 2. Similarities and differences between the principal mammalian single-strand DNA damage repair pathways.** Examples of cell stressors are shown, with resultant DNA damage. DNA repair proceeds through a conserved general mechanism of damage recognition, lesion excision and processing, DNA repair synthesis and ligation of DNA ends, as shown from top to bottom. Components vary between pathways although there is considerable overlap. The four main repair pathways for single-strand DNA lesions are shown, with the key proteins involved. (A) Mismatch repair (MMR). (B) Base excision repair (BER). (C) Nucleotide excision repair (NER). (D) Single-strand break repair (SSBR). AP endonuclease, apurinic/apyrimidinic endonuclease 1; CSA/CSB, Cockayne syndrome protein A/B; ERCC1, excision-repair cross-complementing 1; FEN1, flap endonuclease 1; GG-NER, global genomic nucleotide excision repair; HR23B, human RAD23 homologue B; LIG, DNA ligase; MLH, MutL protein homologue; MSH, MutS protein homologue; PARP, poly(ADP-ribose) polymerase; PMS2, postmeiotic segregation increased 2; PNKP, polynucleotide kinase 3'-phosphatase; Pol, DNA polymerase; TC-NER, transcription-coupled nucleotide excision repair; TFIIH, transcription factor IIH; UV, ultraviolet; XPA/XPC/XPD/XPF/XPG, DNA repair proteins in different xeroderma pigmentosum (XP) complementation groups; XRCC1, X-ray repair cross-complementing 1.

G-quadruplexes and R-loops (Mirkin, 2007; Neil et al., 2017). These structural perturbations of DNA have been implicated in both the normal regulation of cellular functions, such as chromatin organisation and gene expression, and in the aberrant DNA processing that can lead to genomic instability. Repetitive genomic loci are often polymorphic but cells have homeostatic mechanisms to maintain fairly stable repeat lengths in DNA based on structural stability, protein binding and reaction kinetics (Hall et al., 2017; Lee and McMurray, 2014). However, sometimes these mechanisms fail and repeats expand or contract significantly, often with resulting pathology.

Trinucleotide repeat disorders are human diseases that are defined by expanded tandem arrays of three-nucleotide units in the transcribed regions of a diverse range of genes (Budworth and McMurray, 2013). Strikingly, all of these diseases have at least some element of neurological dysfunction suggestive of a specific need to regulate trinucleotide repeats tightly in the nervous system (Orr and Zoghbi, 2007). Diseases caused by expanded tandem CAG repeats in exons constitute a subset of the broader trinucleotide repeat disorder group, and are linked by central neurodegeneration, as well as DNA sequence. They are considered in more detail below.

The first disease shown to be caused by an expanded CAG repeat was spinal and bulbar muscular atrophy (SBMA) in 1991 (Spada et al., 1991). Shortly afterwards, various other dominantly inherited neurodegenerative conditions were also linked to expanded CAG tracts, including HD, dentatorubral-pallidolusian atrophy (DRPLA), and some spinocerebellar ataxias (SCA1,2,3,6,7,12,17) (Koide et al., 1994; Nagafuchi et al., 1994; Orr et al., 1993; The Huntington’s Disease Collaborative Research Group, 1993). Although these diseases all share a common underlying mutation type, genotype-phenotype relationships are not straightforward. Expanded CAG repeats can variably cause autosomal dominant ataxia (SCAs), chorea (HD or DRPLA), or neither (SBMA), in association with a variety of other symptoms. Only a small subset of SCAs are caused by CAG repeats; of the 43 autosomal dominant SCAs described to date, just seven have been attributed to CAG repeats (Synofzik and Schüle, 2017). In addition, there is considerable phenotypic diversity between even those SCAs caused by CAG repeat expansion: for example, SCA6 presents as a fairly ‘pure’ ataxia, whereas SCA7 often has associated retinal degeneration (Sun et al., 2016; Synofzik and Schüle, 2017).

However, even though spinocerebellar degeneration can be caused by a wide variety of CAG repeat and non-CAG mutations in disparate genes, there might be a shared molecular pathogenesis. In support of this hypothesis, many SCA gene products interact with, and presumably mediate their effects through, a limited set of intracellular proteins, the ‘ataxia interactome’ (Kahle et al., 2011; Lim et al., 2006). Pathogenic CAG expansions are also found in some patients with autosomal dominant chorea as part of their clinical presentation; for example, in HD (with psychiatric, behavioural and cognitive symptoms) (Bates et al., 2015) or in DRPLA (with myoclonic epilepsy and ataxia) (Wardle et al., 2009). Again, partially shared clinical phenotypes and underlying mutations hint at a common pathology, and there might be a broader commonality with the CAG repeat SCAs, as exemplified by SCA17, which is also known as Huntington’s disease-like 4 (HDL-4), owing to its clinical presentation (Gövert and Schneider, 2013).

Pathogenic CAG repeat expansions are found in the exons of different genes for each of the different diseases (Table 1). Wild-type repeat numbers range from 4 to ~40 and are polymorphic at each locus. When transcribed and translated, a tandem CAG repeat tract, (CAG)<sub>n</sub>, encodes a polyglutamine stretch in protein, and this polypeptide could have important effects in cells outwith the function of the endogenous protein in which it sits (Ashkenazi et al., 2017; Fujita et al., 2013). Expansion of the tandem CAG repeat over a threshold is necessary and sufficient for all the CAG repeat diseases, suggesting a toxic gain of function. The toxic threshold is usually of the order of 35-45 tandem repeats, although with some variation; for example, the disease threshold is shorter in SCA6 (>19 repeats) and longer in SCA3 (>60 repeats) (Table 1) (Durr, 2010). Once over the disease threshold, longer repeat lengths are associated with earlier symptom onset, although there is considerable variation (Bates et al., 2015; Durr, 2010). Toxicity is conferred by the expanded repeat, but the mechanisms by which CAG repeat expansion leads to specific neurodegeneration remain incompletely understood. Theoretically, CAG-containing DNA and/or mRNA and/or polyglutamine-containing proteins could be pathogenic. Repetitive DNA elements affect gene expression (Gymrek et al., 2015), and there is also evidence of the toxicity of both CAG-containing mRNA and polyglutamine in cells (Cheng et al., 2015; Rué et al., 2016). Additionally, recent evidence from human HD brains has suggested that sense and antisense mRNAs from (CAG)<sub>n</sub> can be translated in all possible frames by noncanonical repeat-

**Table 1. Genomic characteristics of CAG repeat diseases\***

Disease	Gene	Genomic locus	Exon	WT repeat number	Pathogenic CAG repeat threshold <sup>a</sup>	Repeat interruptions?	CAG repeat structure
HD	<i>HTT</i>	4p16	1 of 67	9-26	>35	Yes	(CAG) <sub>n</sub> (CAA) <sub>0-1</sub> CAG
DRPLA	<i>ATN1</i>	12p13	5 of 10	6-35	>47	Not reported	
SBMA	<i>AR</i>	Xq12	1 of 8	9-34	>37	Not reported	
SCA1	<i>ATXN1</i>	6p22	8 of 9	6-35	>38	Yes	(CAG) <sub>n</sub> (CAGCAT) <sub>0-4</sub> (CAG) <sub>n</sub>
SCA2	<i>ATXN2</i>	12q24	1 of 25	13-31	>32	Yes	[(CAG) <sub>n</sub> (CAA) <sub>0-1</sub> (CAG) <sub>n</sub> ] <sub>1-4</sub>
SCA3	<i>ATXN3</i>	14q32	8 of 11	11-44	>60	Yes	(CAG) <sub>2</sub> CAAAAGCAGCAA(CAG) <sub>n</sub>
SCA6	<i>CACNA1A</i>	19p13	47 of 47	4-18	>18	Not reported	
SCA7	<i>ATXN7</i>	3p14	3 of 13	4-19	>33	Not reported	
SCA12	<i>PPP2R2B</i>	5q32	7 of 16 <sup>b</sup>	4-32	>42	Not reported	
SCA17	<i>TBP</i>	6q27	2 of 8	25-40	>40	Yes	(CAG) <sub>3</sub> (CAA) <sub>3</sub> (CAG) <sub>n</sub> CAA CAGCAA(CAG) <sub>n</sub> CAACAG

\*Information from www.genereviews.org; www.scabase.eu; McMurray, 2010; Sequeiros et al., 2010.

<sup>a</sup>Based on uninterrupted CAG repeat tracts and including incomplete penetrance alleles.

<sup>b</sup>5’ UTR of alternatively spliced transcripts.

AR, androgen receptor; *ATN1*, atrophin 1; *ATXN*, ataxin; *CACNA1A*, calcium channel (voltage-dependent), alpha 1A subunit; DRPLA, dentatorubral-pallidolusian atrophy; HD, Huntington’s disease; *HTT*, huntingtin; *PPP2R2B*, protein phosphatase 2 regulatory subunit beta; SBMA, spinal and bulbar muscular atrophy; SCA, spinocerebellar ataxia; *TBP*, TATA-box binding protein.

associated non-ATG (RAN) translation to produce toxic homopolymeric proteins (see Glossary, Box 1) (Bañez-Coronel et al., 2015). Although all underpinned by similar CAG repeat expansions, the different disease phenotypes are associated with the selective degeneration of different brain cell types; for example, cerebellar Purkinje cells are affected in SCAs and striatal medium spiny neurons (MSNs; see Glossary, Box 1) in HD. The reasons for this selectivity are unclear, but emphasise how gene expression, protein context and cell type can all influence CAG repeat pathology.

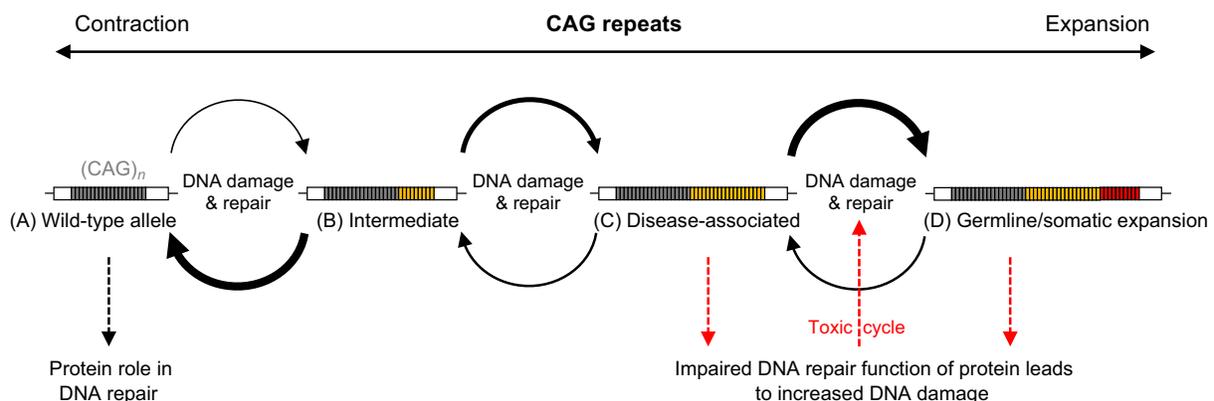
Given their shared causative repeat expansions and overlapping clinical phenotypes, the CAG repeat disorders might be linked by a common pathogenesis at the DNA level, involving DNA damage and repair in neurons. We discuss the intersection of CAG repeat disorders with DNA repair in more detail below.

### CAG repeat disorders and DNA repair

Links between DNA repair defects and neurodegenerative diseases have been known for many years. Fibroblasts and lymphocytes cultured from patients with HD, Alzheimer's disease, Parkinson's disease and amyotrophic lateral sclerosis have all been shown to be sensitive to DNA damage induced by ionising radiation or exogenous chemical mutagens (Moshell et al., 1980; Robison and Bradley, 1984; Scudiero et al., 1981). It has been suggested that accumulation of DNA damage as a result of inadequate DNA repair could cause neurodegeneration, although it has been difficult to discriminate between this hypothesis and the accrual of DNA damage caused by other pathological cellular dysfunction (Robison and Bradley, 1984). The discovery of neurodegenerative CAG repeat disorders, and the apparent similarity of their repeat length variation to that observed in microsatellites of some colorectal cancers, led to a second line of investigation: the role of DNA repair in the modulation of CAG repeat length. However, microsatellite instability is observed throughout the genome in the MMR-deficient tumours of Lynch syndrome, a cancer-predisposition disorder (see Glossary, Box 1), alongside a globally elevated mutation rate. By contrast, HD/SCA patients only seem to show repeat number variation at a disease-specific,

expanded CAG repeat locus (Goellner et al., 1997; Slean et al., 2008). These patients also have a significantly reduced incidence of cancer [e.g. a standardised incidence ratio of 0.47 in the largest study of HD (Ji et al., 2012)].

Once the disease-causing threshold is crossed, CAG repeat length has an inverse relationship with the age at symptom onset. However, in HD, the most well-studied CAG repeat disorder, repeat length only explains ~50% of the observed variation in age at symptom onset. Studies of the large Venezuelan HD kindreds indicated that as much as 40% of the remaining variation was heritable, suggesting that background genetic variants can have a large influence on when symptoms start (Wexler et al., 2004). In order to identify these genetic modifiers (see Glossary, Box 1), a genome-wide association study (GWAS; see Glossary, Box 1) was recently performed using data from just over 4000 HD patients, to look for loci associated with earlier or later onset HD than predicted by CAG repeat length alone (GeM-HD Consortium, 2015). This study identified variants at a number of loci with significant associations with the age at symptom onset. Many of these variants are in, or near, the genes that encode components of DDR pathways, and particularly those involved with MMR (GeM-HD Consortium, 2015). Subsequent work showed that many of the same genetic modifiers are significant in other CAG repeat disorders, suggesting that there is a common pathogenic mechanism driving disease onset, perhaps at the level of the somatic CAG repeat (Bettencourt et al., 2016). Furthermore, a comparison of disease progression with genotype in a sample of HD patients that had been phenotyped in detail showed a genome-wide significant signal in *MSH3*, a MMR gene (Hensman-Moss et al., 2017). Collectively, these results were the first to link human CAG repeat disorder phenotypes directly to DNA repair, and corroborated many earlier results from model systems. The simplest explanation for the genetic data is that DNA repair variants directly affect repeat number in individuals, but it is also possible that expanded (CAG)<sub>n</sub> could exacerbate DNA repair defects (Fig. 3). In the sections below, we explore the evidence that links CAG repeat diseases and DNA repair either at the level of the CAG repeat in the genome, or as a downstream consequence of an expanded CAG repeat.



**Fig. 3. DNA damage and repair can affect CAG repeat length with downstream effects on disease pathogenesis.** CAG repeats in DNA are unstable, and cycles of DNA damage and repair can lead to changes in repeat length. (A) Wild-type length repeats can expand to (B) intermediate lengths, stochastically. These will mostly be repaired to wild-type length (black, bold arrow from B to A), perhaps through a dedicated pathway, but a small number will expand further (C) into the disease-associated range in gametes. (D) Once over the disease-causing threshold, repeats are predisposed to expand further (black, bold arrow from C to D) in both germline and somatic cells. In addition to the role of DNA repair in repeat length changes, some genes containing CAG repeats encode proteins with roles in DNA repair. Expanded repeats can impair the functions of these DNA repair proteins, leading to the accrual of DNA damage in neurons and to a toxic cycle of DNA damage/repair and repeat expansion (red dashed arrows). DNA repair variants associated with earlier or later disease onset could affect any of these processes.

### CAG repeats in the genome

Expansion of a tandem CAG repeat in genomic DNA over a threshold number is absolutely required for each of the CAG repeat disorders (Table 1). Long repeats are intrinsically unstable, as shown in cell-free and unicellular systems, and their dynamics in neurons and gametes are linked to disease pathology.

#### Intrinsic instability of CAG repeat number

Biophysical studies *in vitro* have shown that disease-causing CAG repeats can form unusual DNA structures, including stable hairpins. The stability of these structures correlates with the propensity for CAG repeat expansion (Gacy et al., 1995). CAG:GAC base pairing in the stem of a hairpin contains a middle A:A base pair mismatch, which *in silico* modelling predicts will adopt an unusual Z DNA structure, perhaps via the flipping out of bases (Khan et al., 2015). This could predispose hairpin structures to both increased DNA damage and MMR protein binding. Indeed, bases damaged by ROS (e.g. 8-oxo-dG) can affect the formation and stability of hairpins, as well as having consequences on DNA repair fidelity (Volle et al., 2012). MSH2, part of the MutS mismatch recognition complexes (see Glossary, Box 1), binds directly to slipped-strand DNA structures formed by (CAG)<sub>n</sub> *in vitro* (Pearson et al., 1997). The processing of these artificial DNA substrates by human cell extracts requires various repair factors, principally MMR proteins (including MSH2, MSH3, PMS2), although the repair outcomes depend on the starting DNA structure and not just its sequence (Panigrahi et al., 2005, 2010). It is hard to draw physiological mechanistic conclusions from these cell-free systems, but putative pathways can be identified; for example, MutSβ (MSH2/MSH3 complex) is required for repeat expansion in some assays (Nakatani et al., 2015; Stevens et al., 2013), and long-patch BER deficiency has been implicated elsewhere (Goula et al., 2009).

Further insight into CAG repeat stability has come from bacteria and yeast. These microorganisms are attractive models as they are genetically tractable, most of their DNA repair factors have been identified, and high-throughput screening assays exist for them (Bichara et al., 2006; Dixon et al., 2004; Kim et al., 2016). Repeats are unstable in these microorganisms, and seem to have a similar length threshold to that observed in human diseases, although they exhibit a propensity for repeat contractions over expansions. There is also a requirement for MMR in (CAG)<sub>n</sub> instability: when MMR factors are knocked out, repeats are stabilised (Jaworski et al., 1995; Williams and Surtees, 2015). Given the phylogenetic conservation of MMR and other DNA repair factors from bacteria to humans (Fishel, 2015), these results might be relevant to human disease. However, bacterial and yeast cells divide in culture, and have much simpler DNA-damage response systems than human cells. They also contain eukaryotic triplet repeats out of genomic context, either on plasmids or integrated into DNA that lacks human chromatin structure and organisation. More disease-relevant data have come from multicellular organisms, as detailed below.

#### Germline and somatic instability of CAG repeats

Heritable, stochastic CAG repeat expansions in germline cells (i.e. sperm or egg) are reported in many CAG and non-CAG repeat disorders (Budworth and McMurray, 2013; Durr, 2010). Given that repeat length correlates inversely with age at disease onset, this intergenerational propensity for CAG expansion explains the phenotypic observation of anticipation, whereby disease onset tends to get earlier over generations. This has been shown in all CAG repeat disorders, and is usually more marked through the paternal line (Durr, 2010), although other non-CAG trinucleotide

repeat disorders, such as Huntington's disease-like 2 (HDL2) and myotonic dystrophy, show increased anticipation through the maternal line (Gövert and Schneider, 2013). Repeat length analysis has shown that significant CAG mosaicism is present in the sperm of male HD patients and that this correlates with repeat expansion on transmission (Telenius et al., 1995). Evidence from a transgenic mouse model of HD (R6/1) suggests that repeat expansion in sperm occurs after meiosis, as haploid spermatids are maturing into spermatozoa (Kovtun and McMurray, 2001). This implicates DNA repair rather than DNA replication in the DNA synthesis needed for repeat expansion.

Variations in CAG repeat length have also been found in somatic (i.e. nongermline) cells. Repeat length stability varies across cell types and can be associated with phenotype. For example, in transgenic and knock-in mouse models of HD, CAG repeat length tends to be increased in cells from the striatum, cortex and liver, but stable in cells from the cerebellum, blood and tail. Maximum expansion is observed in the striatum, which correlates well with the degeneration of striatal MSNs that underpins the disease (Gonitel et al., 2008; Lee et al., 2011; Møllersen et al., 2010). Large CAG repeat expansions have also been demonstrated in postmortem human brain neurons from HD patients, and increased expansion of repeats correlates with younger age at onset of symptoms (Kennedy, 2003; Shelbourne et al., 2007; Swami et al., 2009). DNA repair is implicated in affected neurons, as these are postmitotic. However, similar tissue-specific patterns of repeat expansion are seen in SCAs and, in these diseases, degeneration is observed in the cerebellar Purkinje neurons rather than the striatal MSNs (Chong et al., 1995; Hashida et al., 1997; Tanaka et al., 1996). The relatively high levels of expression of certain DNA repair factors in the cerebellum might prevent significant repeat expansion in this tissue. Genomic context is also important. For example, SCA7 CAG repeats in a transgenic mouse model were stable when present in complementary DNA (cDNA), but unstable if contained in a human genomic fragment. In addition, repeat stability did not correlate with neurodegeneration (Libby et al., 2003). Therefore, although somatic CAG repeat expansion can occur, and might correlate with neurodegeneration in HD, it remains unclear whether expansion drives pathology in individuals or results from downstream DNA repair defects in affected cells (Chong et al., 1995).

#### Modifiers of CAG repeat stability in the genome

Various *cis* and *trans* factors can affect the stability of CAG repeats in the genome. *Cis* factors include the length of the CAG repeat, the presence of repeat interruptions, haplotype and genomic context. Longer repeats, particularly those that are over the disease-causing threshold, are more unstable and tend to expand (Fig. 3), although repeat interruptions can temper these effects (Usdin et al., 2015). Interruptions have been identified and associated with increased repeat stability, and later age at disease onset, in various SCAs and in HD (Table 1), as well as in other trinucleotide repeat disorders such as fragile X syndrome, Friedreich's ataxia and myotonic dystrophy (Chong et al., 1997; Chung et al., 1993; Eichler et al., 1994; Gao et al., 2008; Yu et al., 2011). Within the CAG repeat diseases, interruptions of a pure CAG repeat almost always involve alteration of the third base in the codon (the most tolerant of change). Interruptions were first observed in SCA1 with variable numbers of CAT codons replacing CAG (Chung et al., 1993). These reduce the stability of hairpin structures formed by pure (CAG)<sub>n</sub> repeats in DNA, inhibiting repeat expansion, as well as inserting a number of histidines into the polyglutamine stretch of protein

(Menon et al., 2013; Pearson et al., 1998; Sobczak and Krzyzosiak, 2004). More commonly, CAG repeat tracts can be interrupted by CAA codons, which alter the DNA/RNA sequence but do not affect the translated polyglutamine. In SCA2, pure CAG repeats of intermediate length (~27-35 repeats) can expand into the disease range but repeats interrupted by CAA expand to a lesser extent and are associated with different neurological phenotypes (amyotrophic lateral sclerosis or parkinsonism, depending on the repeat number) (Charles et al., 2007; Elden et al., 2010; Yu et al., 2011). In SCA17, complex CAG repeat alleles of the TATA-box binding protein gene (*TBP*) that contain various CAA interruptions are associated with later disease onset (Gao et al., 2008). In HD, the penultimate codon of the CAG tract is usually CAA, followed by one more CAG before the  $(CCG)_n$  repeat encoding polyproline begins. The relevance of the penultimate interrupting CAA is uncertain, but its absence correlates with earlier onset (and with repeat expansion) in a few cases (Goldberg et al., 1995). Conversely, in a bacterial artificial chromosome mouse model of HD, a huntingtin (*HTT*) allele with 97 codons of alternating CAA/CAG is stable over 12 months in both germline and somatic cells, but the mice still develop a neurodegenerative pathology (Gray et al., 2008).

It appears that if a CAG repeat is above a threshold length and of appropriate codon structure and context, it is licensed to become unstable. *Trans* factors, such as DNA repair proteins, can then act on this DNA substrate to increase or decrease the repeat number in specific tissues. Work on *trans* factors has mostly been carried out in mouse models of HD. Transgenic and knock-in mouse models of HD, all with long tandem CAG tracts of >100 repeats, develop progressive neurological impairment, leading to reduced abilities in tests of motor, coordination and cognitive function (Brooks et al., 2012; Ferrante, 2009). The development of somatic CAG repeat expansion in the striatum of mouse HD models (for example, R6/1 transgenic mice carrying exon 1 of the human *HTT* gene) (Mangiarini et al., 1997) correlates with symptom development. Crossing mouse models of HD with mice that carry different DNA repair gene mutations has shown that deficiencies in specific MMR genes (e.g. *Msh2*, *Msh3*, *Mlh1* or *Mlh3*), or BER genes (e.g. *Ogg1* or *Nei1*) can abrogate somatic and/or germline CAG repeat expansion and, in some cases, ameliorate HD-like phenotypes (Budworth et al., 2015; Kovalenko et al., 2012; Pinto et al., 2013; Usdin et al., 2015; Wheeler et al., 2003). These effects seem gene specific, as knockouts of other DNA repair factors in the same pathways [e.g. of *Msh6* (MMR) or *Mpg* glycosylase (BER)] have no effect on repeat stability. In addition, results from different diseases and models have been inconsistent. For example, NER factors have been implicated in CAG repeat diseases. A study of 137 SCA3 parent-child repeat transmissions identified variants in NER factors Cockayne syndrome protein B (CSB; see Glossary, Box 1; Fig. 2C), RPA proteins and CDK7, as being associated with intergenerational repeat expansions (Martins et al., 2014). However, knockouts of different NER factors in different CAG repeat disease models have variable effects: *Csb* (*Ercc6*) knockout in HD mice promotes germline repeat instability (Kovtun et al., 2011); *XPG* (*mus201*) knockout in a *Drosophila* model of SCA3 abolishes repeat instability (Jung and Bonini, 2007); *Xpa* knockout in a SCA1 mouse reduces somatic repeat instability in many areas of the brain (although not in the cerebellum) (Hubert et al., 2011). The reasons for these differences are not fully understood, although the human homologues might not contain relevant variation, and some repair factors have additional functions outside NER; for example, CSB has chromatin remodelling and transcriptional regulation activities.

### Downstream deficits in DNA repair in CAG repeat diseases

The proteins encoded by CAG repeat-containing genes have a wide range of functions in different cellular processes (Table 2). However, most are ubiquitously expressed and have links to transcriptional regulation, tying in with earlier work that identified polyglutamine stretches of 10-30 amino acids as being potent transactivators (Gerber et al., 1994). The recent realisation that repeat length polymorphism in DNA can modulate gene expression suggests that expanded repeats might affect transcription through both DNA- and protein-mediated mechanisms (Gymrek et al., 2015). The target genes that are differentially expressed as the result of repeat instability are not known. Fibroblasts, lymphoblasts or lymphocytes cultured from patients with HD or other neurodegenerative conditions accrue more DNA damage than wild-type controls; this could be caused by defective DNA repair (Moshell et al., 1980; Robison and Bradley, 1984; Scudiero et al., 1981). In support of this hypothesis, various polyglutamine-containing proteins, including HTT, androgen receptor (AR), ataxin (ATXN) 1, ATXN2 and ATXN3, have been shown to have roles in DNA repair (Table 2), and there is evidence that at least some of the pathogenesis of CAG repeat expansion might arise from the loss of wild-type protein function (Ashkenazi et al., 2017; Chatterjee et al., 2015; Gao et al., 2015; Saudou and Humbert, 2016; Zeitlin et al., 2000). As an example, regulated phosphorylation of HTT at various sites is involved in the DNA damage response. The N-terminus of HTT (particularly methionine 8) can act as a direct sensor of oxidative stress, leading to its phosphorylation at serines 13 and 16 and translocation from its usual cytoplasmic location to the nucleus (DiGiovanni et al., 2016). In the nucleus, HTT is recruited to sites of DNA damage, in a process dependent on ATM serine/threonine kinase (ATM), and might function as a scaffold for DNA repair complexes (Fig. 4A). Analysis of patient fibroblasts shows that an expanded polyglutamine in HTT does not prevent its recruitment to sites of DNA damage, but is associated with increased DNA damage, consistent with a dominant-negative effect on repair (Fig. 4B) (Maiuri et al., 2016). In addition, wild-type HTT is phosphorylated by cyclin-dependent kinase (CDK) 5 on serines 1181 and 1201 in response to DNA damage, and this seems to have a protective role in inhibition of p53 (TP53)-induced cell death. Loss of this protective phosphorylation in ageing and/or disease-associated neurons could lead to increased cell death, although the exact molecular mechanisms involved are not understood (Fig. 4B) (Anne et al., 2007).

Recent work on ATXN3 has reinforced the importance of subcellular localisation and protein-protein interactions in the regulation of DNA repair. ATXN3 is a deubiquitinase that has a neuroprotective role mediated by its regulation of autophagy (via wild-type polyglutamine; see Glossary, Box 1) (Ashkenazi et al., 2017), the ubiquitin-proteasome system, and DNA repair. Heat-shock or oxidative stress can stimulate the nuclear translocation of ATXN3, where it has roles in transcriptional regulation, possibly through its deubiquitinase activity, and DNA repair (Fig. 4A) (Orr, 2012). The latter was first suggested by the interaction of ATXN3 with human RAD23 homologues in a yeast two-hybrid screen, as RAD23 proteins are involved in NER (Fig. 2C) (Wang et al., 2000). More recently, ATXN3 has been shown to protect cells from DNA damage through its interaction with polynucleotide kinase 3'-phosphatase (PNKP), a key enzyme involved in the processing and repair of DNA strand breaks (Figs 1 and 2) (Chatterjee et al., 2015). Rare mutations in PNKP can cause ataxic syndromes (Bras et al., 2015). Interestingly, the expansion of a polyglutamine tract in

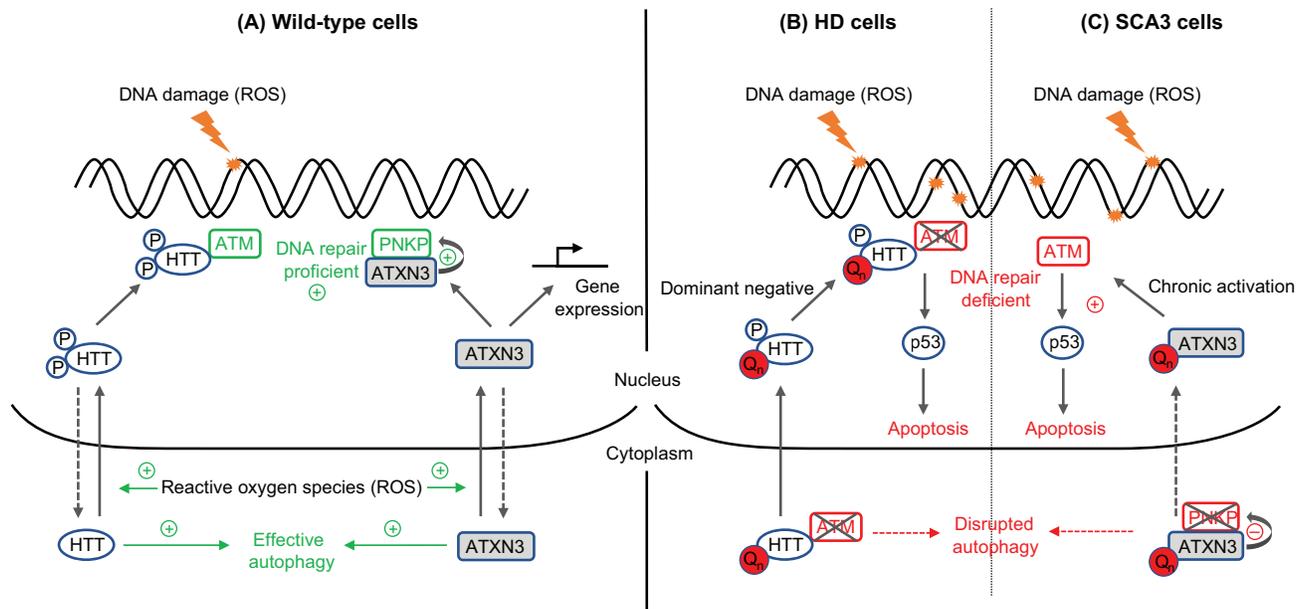
**Table 2. Functions of proteins encoded by genes causing CAG repeat diseases, and their links to DNA damage and repair**

Disease	Gene	Wild-type protein functions <sup>a</sup>	Expression <sup>b</sup>	Links to DNA damage/repair	References
HD	<i>HTT</i>	Transcriptional regulation; molecular scaffolding and vesicle trafficking; neurodevelopment; cell survival	Ubiquitous	N-terminus (M8) functions as a ROS sensor leading to nuclear translocation DNA damage leads to phosphorylation (serines 1181 and 1201) by Cdk5 as part of DDR Recruited by ATM to sites of DNA damage Quiescent human HD fibroblasts are defective in DSB repair. Mutant HTT may sequester ATM in cytoplasm Exaggerated DDR following oxidative stress in HD fibroblasts	DiGiovanni et al., 2016 Anne et al., 2007 Maiuri et al., 2016 Ferlazzo et al., 2014 Giuliano et al., 2003
DRPLA	<i>ATN1</i>	Transcriptional co-repressor through recruiting NR2E1	Ubiquitous	None known	
SBMA	<i>AR</i>	Transcription factor when bound to androgen	Testis, breast, liver, platelets. Low levels elsewhere	AR with expanded polyglutamine can sequester PTIP (containing glutamine-rich region) away from DNA repair pathways, leading to accumulation of DNA damage in cell models	Xiao et al., 2012
SCA1	<i>ATXN1</i>	Brain development via transcriptional co-repressor complex with capicua protein; alternative splicing; cell signalling through Notch; modulation of PP2A	Ubiquitous	Polyglutamine-containing ATXN1 (or HTT, AR, ATXN7) can sequester multifunctional VCP, leading to functional deficiency in DNA repair and accumulation of DNA damage in cells Overexpression of DNA repair factor RPA1 in mouse or <i>Drosophila</i> models of SCA1 can ameliorate phenotype	Fujita et al., 2013 Barclay et al., 2014; Taniguchi et al., 2016
SCA2	<i>ATXN2</i>	RNA metabolism; regulation of translation	Ubiquitous	shRNA knockdown of ATXN2 in HeLa cells leads to increased DNA damage (DSBs and R-loops), partially rescued by Mg <sup>2+</sup> supplementation Exaggerated DDR following oxidative stress in SCA2 fibroblasts	Abraham et al., 2016 Giuliano et al., 2003
SCA3	<i>ATXN3</i>	Transcriptional regulation (stress response); protein homeostasis through ubiquitin-proteasome system (ataxin-3 is a deubiquitinase)	Ubiquitous	RAD23A/B have roles in NER and proteasome function. They bind ATXN3 and protect it from proteasomal degradation ATXN3 with expanded polyglutamine sequesters PNKP outside nucleus and inhibits its 3'-phosphatase activity, leading to increased DNA strand breaks in cell and mouse models, and postmortem human brains	Blount et al., 2014 Chatterjee et al., 2015; Gao et al., 2015
SCA6	<i>CACNA1A</i>	Voltage-gated calcium channel abundant in cerebellar Purkinje cells; product of alternative translation functions as a transcription factor involved in neuronal differentiation	Predominantly neuronal	None known	Du et al., 2013
SCA7	<i>ATXN7</i>	Component of STAGA chromatin remodelling complex that regulates transcription	Ubiquitous	None known	Wang and Dent, 2014
SCA12	<i>PPP2R2B</i>	Regulatory subunit B of PP2A involved in transcriptional regulation, cell growth and division	Predominantly neuronal	None known	Cohen and Margolis, 2016
SCA17	<i>TBP</i>	Binds TATA box in gene promoters as part of TFIID, which is required for initiation of transcription by RNA polymerase II	Ubiquitous	TBP can bind damaged DNA at or near TATA boxes	Aboussekhra and Thoma, 1999; Jung et al., 2001

<sup>a</sup>Wild-type protein functions from www.genecards.org and specific references: HD (Liu and Zeitlin, 2017; Saudou and Humbert, 2016); SCA1 (Lu et al., 2017; Sánchez et al., 2016); SCA2 (Ostrowski et al., 2017); all SCAs (Orr, 2012).

<sup>b</sup>Protein expression from www.proteinatlas.org and www.genecards.org (Human Integrated Protein Expression Database).

AR, androgen receptor; ATM, ataxia telangiectasia mutated; *ATN1*, atrophin 1; *ATXN*, ataxin; *CACNA1A*, calcium channel (voltage-dependent), alpha 1A subunit; DDR, DNA damage response; DRPLA, dentatorubral-pallidolysian atrophy; DSB, double-strand DNA break; HD, Huntington's disease; *HTT*, huntingtin; NER, nucleotide excision repair; PNKP, polynucleotide kinase 3'-phosphatase; STAGA, SPT3-TAF9-GCN5 acetyltransferase; PP2A, protein phosphatase 2; *PPP2R2B*, protein phosphatase 2 regulatory subunit beta; PTIP, PAX2 transactivation domain-interacting protein; ROS, reactive oxygen species; SBMA, spinal and bulbar muscular atrophy; SCA, spinocerebellar ataxia; shRNA, short hairpin RNA; TBP, TATA-box binding protein; TFIID, transcription factor IID; VCP, valosin-containing protein.



**Fig. 4. Putative roles of HTT and ATXN3 in DNA repair and how HTT and ATXN3 polyglutamine expansions might lead to DNA damage and apoptosis.** (A) Wild-type HTT and ATXN3 proteins have various roles in the DNA damage response, as illustrated. ROS that cause DNA damage also induce the nuclear translocation of both HTT and ATXN3 (solid arrows), as well as specific HTT phosphorylation. In the nucleus, HTT is recruited to sites of DNA damage by ATM, and can act as a scaffold for DNA repair processes. Nuclear ATXN3 can bind to and stimulate the DNA end-processing repair factor, PNKP, as well as altering gene expression as part of the cellular stress response. HTT and ATXN3 also have functions in regulation of autophagy. Repair processes and their associated factors are shown in green. (B) In HD or (C) in SCA3, disease-length polyglutamine expansions (depicted as red, Q<sub>n</sub>) can inhibit DNA repair processes, leading to the accrual of DNA damage in cells. In both of these diseases, the mutated proteins can sequester DNA repair factors in the cytoplasm (ATM in HD, PNKP in SCA3), away from sites of DNA damage. Persistent DNA damage and signalling can result in p53-mediated apoptosis; in HD, via dominant-negative hypophosphorylated mutant HTT at sites of DNA damage; in SCA3 via chronic activation of ATM by mutant ATXN3. Nonfunctional ATM and PNKP are crossed in the figure. ATM, ataxia telangiectasia mutated; ATXN3, ataxin-3; HD, Huntington's disease; HTT, huntingtin; PNKP, polynucleotide kinase 3'-phosphatase; ROS, reactive oxygen species; SCA3, spinocerebellar ataxia type 3.

ATXN3 leads to both the sequestration of PNKP outside the nucleus and the inhibition of nuclear PNKP activity, which together lead to impaired DNA repair. Persistent DNA damage (and DNA damage signalling through ATM) is observed in mouse and cellular models of SCA3, as well as in postmortem human SCA3 brains, and this can trigger cell death through p53-mediated and other pathways (Fig. 4C) (Chatterjee et al., 2015; Gao et al., 2015).

A similar theme is continued in cell models of SBMA and SCA1. Expanded polyglutamine within AR or ATXN1, respectively, lead to the sequestration of proteins involved in DNA repair and the subsequent accumulation of DNA damage (Fujita et al., 2013; Xiao et al., 2012). As well as being intrinsically deleterious to cells, unrepaired DNA damage also leads to persistent activation of the DDR, which in itself can be toxic, as seen in other neurodegenerative disorders (Hoch et al., 2016). The importance of accurate and timely DNA repair in the cerebellum has long been noted, given that mutations in enzymes such as aprataxin and tyrosyl DNA phosphodiesterase 1 (TDP1), responsible for processing damaged DNA ends to permit repair (Fig. 1), lead specifically to cerebellar neurodegeneration (Ward and La Spada, 2015).

Therefore, there is increasing evidence in a range of CAG repeat diseases that defective DNA repair might be involved in disease pathogenesis. An expanded polyglutamine tract could lead to the inactivation and/or inappropriate sequestration of repair proteins, such that DNA damage builds up in neurons. If the wild-type proteins themselves have functions in DNA repair, then pathogenesis could be the result of a combination of loss of function and dominant-negative gain of function (Fig. 4). Accrual of DNA damage would also predispose CAG repeats, which are

already susceptible to damage as discussed above, to further strand breaks, the repair of which could result in further repeat expansion, thus setting up a toxic cycle (Fig. 3C,D).

**Conclusion**

CAG repeat disorders consist of a set of overlapping diseases that are linked by pathogenic repeat expansion, neurodegeneration and lack of disease-modifying therapies. Although some of the causative mutations have been known for 25 years, very little progress has been made in translating findings from cell and animal models of these diseases into new treatments. The recent discovery of disease-modifying genetic variants in HD and SCAs has shown the power of 'natural' clinical trials, which capture information on modifying variants that have been 'crossed' onto disease-causing CAG repeat mutations present in the population (Holmans et al., 2017). Excitingly, many of the identified disease-modifying variants converge on specific DNA repair pathways, such as MMR. Understanding why this is the case could illuminate the link between CAG repeat expansion and neurodegeneration. Variants could be influencing CAG repeat pathogenesis by affecting the CAG repeat itself and/or by modulating the downstream DNA damage that results from a defective polyglutamine-containing protein. Ongoing GWAS and the future whole-exome or genome sequencing of individuals with CAG repeat diseases promise to yield more leads, which will need to be validated in human cell models to gain a greater understanding of the underlying molecular mechanisms involved.

Common variation in DDR pathways has also been associated, via GWAS, with a range of other diseases. In the neuropsychiatric

field, variation in the MMR gene *MLH1* has been associated with autism, schizophrenia and lithium-responsive bipolar disorder (Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium, 2017; Ripke et al., 2014; Song et al., 2016), while the broader DDR has been implicated in frontotemporal dementia (Ferrari et al., 2017). Beyond neuropsychiatry, DDR genes have also been associated with lipid metabolism (*MSH3* and *FAN1* loci) (Weissglas-Volkov et al., 2013), reproductive ageing (Day et al., 2015) and longevity (Shadyab and LaCroix, 2015). Inherited mutations in multiple DDR genes are associated with familial cancers, but there is scant evidence for common DDR variation increasing cancer risk in nonfamilial disease. Possible contributions of the DDR genes to testicular cancer (Litchfield et al., 2015) and in a grouped analysis of lung, ovary, prostate, breast and colorectal cancers (Scarborough et al., 2016) have been identified, but the lack of DDR gene and pathway associations in the many very large cancer GWAS implies that variants in these pathways are, at most, likely to have small effect sizes in these diseases. Therefore, variation in DDR pathways is not specific to CAG repeat disorders and could impact multiple diseases in different ways.

The development of olaparib, a poly(ADP-ribose) polymerase (PARP; see Glossary, Box 1; Fig. 2) inhibitor, as a useful antineoplastic drug has shown that therapeutic manipulation of the DNA damage response is feasible in humans (Brown et al., 2017; Pearl et al., 2015). Cross-pollinating advances made in the cancer field, such as the development of synthetic lethality screens (see Glossary, Box 1) in the DDR, with novel genetic leads identified from studying CAG repeat diseases, could help develop drugs more rapidly. Furthermore, therapeutic leads based on genetic discoveries linked directly to human disease phenotypes are more likely to be translated into effective disease-modifying clinical treatments (Nelson et al., 2015; Plenge et al., 2013). Finally, the build-up of DNA damage in ageing neurons, potentially exacerbated by defective DNA repair processes, could represent a broader paradigm for neurodegenerative pathogenesis, making the findings of CAG repeat disease research more widely applicable.

#### Acknowledgements

T.H.M. gratefully acknowledges support from the Wales clinical academic training scheme, Patrick Berthoud Foundation/Association of British Neurologists, and Medical Research Council.

#### Competing interests

The authors declare no competing or financial interests.

#### Funding

This work was supported by the Medical Research Council (MR/P001629/1).

#### References

- Aboussekhra, A. and Thoma, F. (1999). TATA-binding protein promotes the selective formation of UV-induced (6-4)-photoproducts and modulates DNA repair in the TATA box. *EMBO J.* **18**, 433-443.
- Abraham, K. J., Chan, J. N. Y., Salvi, J. S., Ho, B., Hall, A., Vidya, E., Guo, R., Killackey, S. A., Liu, N., Lee, J. E. et al. (2016). Intersection of calorie restriction and magnesium in the suppression of genome-destabilizing RNA-DNA hybrids. *Nucleic Acids Res.* **44**, 8870-8884.
- Anne, S. L., Saudou, F. and Humbert, S. (2007). Phosphorylation of huntingtin by cyclin-dependent kinase 5 is induced by DNA damage and regulates wild-type and mutant huntingtin toxicity in neurons. *J. Neurosci.* **27**, 7318-7328.
- Ashkenazi, A., Bento, C. F., Ricketts, T., Vicinanza, M., Siddiqi, F., Pavel, M., Squitieri, F., Hardenberg, M. C., Imarisio, S., Menzies, F. M. et al. (2017). Polyglutamine tracts regulate beclin 1-dependent autophagy. *Nature* **545**, 108-111.
- Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium (2017). Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol. Autism* **8**, 21.
- Bañez-Coronel, M., Ayhan, F., Tarabochia, A. D., Zu, T., Perez, B. A., Tusi, S. K., Pletnikova, O., Borchelt, D. R., Ross, C. A., Margolis, R. L. et al. (2015). RAN translation in huntington disease. *Neuron* **88**, 667-677.
- Barclay, S. S., Tamura, T., Ito, H., Fujita, K., Tagawa, K., Shimamura, T., Katsuta, A., Shiwaku, H., Sone, M., Imoto, S. et al. (2014). Systems biology analysis of *Drosophila* in vivo screen data elucidates core networks for DNA damage repair in SCA1. *Hum. Mol. Genet.* **23**, 1345-1364.
- Bates, G. P., Dorsey, R., Gusella, J. F., Hayden, M. R., Kay, C., Leavitt, B. R., Nance, M., Ross, C. A., Scahill, R. I., Wetzel, R. et al. (2015). Huntington disease. *Nat. Rev. Dis. Prim.* **1**, 15005.
- Bettencourt, C., Hensman-Moss, D., Flower, M., Wiethoff, S., Brice, A., Goizet, C., Stevanin, G., Koutsis, G., Karadima, G., Panas, M. et al. (2016). DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases. *Ann. Neurol.* **79**, 983-990.
- Bichara, M., Wagner, J. and Lambert, I. B. (2006). Mechanisms of tandem repeat instability in bacteria. *Mutat. Res. Mol. Mech. Mutagen.* **598**, 144-163.
- Blount, J. R., Tsou, W.-L., Ristic, G., Burr, A. A., Ouyang, M., Galante, H., Scaglione, K. M. and Todi, S. V. (2014). Ubiquitin-binding site 2 of ataxin-3 prevents its proteasomal degradation by interacting with Rad23. *Nat. Commun.* **5**, 4638.
- Bras, J., Alonso, I., Barbot, C., Costa, M. M., Darwent, L., Orme, T., Sequeiros, J., Hardy, J., Coutinho, P. and Guerreiro, R. (2015). Mutations in PNKP cause recessive ataxia with oculomotor apraxia type 4. *Am. J. Hum. Genet.* **96**, 474-479.
- Brooks, S. P., Jones, L. and Dunnett, S. B. (2012). Comparative analysis of pathology and behavioural phenotypes in mouse models of Huntington's disease. *Brain Res. Bull.* **88**, 81-93.
- Brown, J. S., O'Carrigan, B., Jackson, S. P. and Yap, T. A. (2017). Targeting DNA repair in cancer: beyond PARP inhibitors. *Cancer Discov.* **7**, 20-37.
- Budworth, H. and McMurray, C. T. (2013). A brief history of triplet repeat diseases. *Methods Mol. Biol.* **1010**, 3-17.
- Budworth, H., Harris, F. R., Williams, P., Lee, D. Y., Holt, A., Pahnke, J., Szczesny, B., Acevedo-Torres, K., Ayala-Peña, S. and McMurray, C. T. (2015). Suppression of somatic expansion delays the onset of pathophysiology in a mouse model of huntington's disease. *PLoS Genet.* **11**, e1005267.
- Charles, P., Camuzat, A., Benammar, N., Sellal, F., Destée, A., Bonnet, A.-M., Lesage, S., Le Ber, I., Stevanin, G., Dürr, A. et al. (2007). Are interrupted SCA2 CAG repeat expansions responsible for parkinsonism? *Neurology* **69**, 1970-1975.
- Chatterjee, A., Saha, S., Chakraborty, A., Silva-Fernandes, A., Mandal, S. M., Neves-Carvalho, A., Liu, Y., Pandita, R. K., Hegde, M. L., Hegde, P. M. et al. (2015). The role of the mammalian DNA end-processing enzyme polynucleotide kinase 3'-phosphatase in spinocerebellar ataxia type 3 pathogenesis. *PLoS Genet.* **11**, e1004749.
- Cheng, H.-M., Chern, Y., Chen, I.-H., Liu, C.-R., Li, S.-H., Chun, S. J., Rigo, F., Bennett, C. F., Deng, N., Feng, Y. et al. (2015). Effects on murine behavior and lifespan of selectively decreasing expression of mutant huntingtin allele by supt4h knockdown. *PLoS Genet.* **11**, e1005043.
- Chong, S. S., McCall, A. E., Cota, J., Subramony, S. H., Orr, H. T., Hughes, M. R. and Zoghbi, H. Y. (1995). Gametic and somatic tissue-specific heterogeneity of the expanded SCA1 CAG repeat in spinocerebellar ataxia type 1. *Nat. Genet.* **10**, 344-350.
- Chong, S. S., Almqvist, E., Telenius, H., LaTray, L., Nichol, K., Bourdelat-Parks, B., Goldberg, Y. P., Haddad, B. R., Richards, F., Sillence, D. et al. (1997). Contribution of DNA sequence and CAG size to mutation frequencies of intermediate alleles for Huntington disease: evidence from single sperm analyses. *Hum. Mol. Genet.* **6**, 301-309.
- Chung, M., Ranum, L. P. W., Duvick, L. A., Servadio, A., Zoghbi, H. Y. and Orr, H. T. (1993). Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I. *Nat. Genet.* **5**, 254-258.
- Cohen, R. L. and Margolis, R. L. (2016). Spinocerebellar ataxia type 12. *Curr. Opin. Neurol.* **29**, 735-742.
- Cooke, M. S., Evans, M. D., Dizdaroglu, M. and Lunec, J. (2003). Oxidative DNA damage: mechanisms, mutation, and disease. *FASEB J.* **17**, 1195-1214.
- Day, F. R., Ruth, K. S., Thompson, D. J., Lunetta, K. L., Pervjakova, N., Chasman, D. I., Stolk, L., Finucane, H. K., Sulem, P., Bulik-Sullivan, B. et al. (2015). Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nat. Genet.* **47**, 1294-1303.
- de Koning, A. P. J., Gu, W., Castoe, T. A., Batzer, M. A. and Pollock, D. D. (2011). Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* **7**, e1002384.
- DiGiovanni, L. F., Mocle, A. J., Xia, J. and Truant, R. (2016). Huntingtin N17 domain is a reactive oxygen species sensor regulating huntingtin phosphorylation and localization. *Hum. Mol. Genet.* **25**, 3937-3945.
- Dixon, M. J., Bhattacharyya, S. and Lahue, R. S. (2004). Genetic assays for triplet repeat instability in yeast. In *Trinucleotide Repeat Protocols. Methods in Molecular Biology* (ed. Y. Kohwi), pp. 029-046. New Jersey: Humana Press.
- Du, X., Wang, J., Zhu, H., Rinaldo, L., Lamar, K.-M., Palmenberg, A. C., Hansel, C. and Gomez, C. M. (2013). Second cistron in CACNA1A gene encodes a

- transcription factor mediating cerebellar development and SCA6. *Cell* **154**, 118-133.
- Durr, A. (2010). Autosomal dominant cerebellar ataxias: polyglutamine expansions and beyond. *Lancet. Neurol.* **9**, 885-894.
- Eichler, E. E., Holden, J. J., Popovich, B. W., Reiss, A. L., Snow, K., Thibodeau, S. N., Richards, C. S., Ward, P. A. and Nelson, D. L. (1994). Length of uninterrupted CGG repeats determines instability in the FMR1 gene. *Nat. Genet.* **8**, 88-94.
- Eiden, A. C., Kim, H.-J., Hart, M. P., Chen-Plotkin, A. S., Johnson, B. S., Fang, X., Armakola, M., Geser, F., Greene, R., Lu, M. M. et al. (2010). Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature* **466**, 1069-1075.
- Evans, S. J. W., Douglas, I., Rawlins, M. D., Wexler, N. S., Tabrizi, S. J. and Smeeth, L. (2013). Prevalence of adult Huntington's disease in the UK based on diagnoses recorded in general practice records. *J. Neurol. Neurosurg. Psychiatry* **84**, 1156-1160.
- Ferlazzo, M. L., Sonzogni, L., Granzotto, A., Bodgi, L., Lartin, O., Devic, C., Vogin, G., Pereira, S. and Foray, N. (2014). Mutations of the huntington's disease protein impact on the ATM-dependent signaling and repair pathways of the radiation-induced DNA double-strand breaks: corrective effect of statins and bisphosphonates. *Mol. Neurobiol.* **49**, 1200-1211.
- Ferrante, R. J. (2009). Mouse models of Huntington's disease and methodological considerations for therapeutic trials. *Biochim. Biophys. Acta Mol. Basis Dis.* **1792**, 506-520.
- Ferrari, R., Lovering, R. C., Hardy, J., Lewis, P. A. and Manzoni, C. (2017). Weighted protein interaction network analysis of frontotemporal dementia. *J. Proteome Res.* **16**, 999-1013.
- Fishel, R. (2015). Mismatch repair. *J. Biol. Chem.* **290**, 26395-26403.
- Fisher, E. R. and Hayden, M. R. (2014). Multisource ascertainment of Huntington disease in Canada: prevalence and population at risk. *Mov. Disord.* **29**, 105-114.
- Fujita, K., Nakamura, Y., Oka, T., Ito, H., Tamura, T., Tagawa, K., Sasabe, T., Katsuta, A., Motoki, K., Shiwaku, H. et al. (2013). A functional deficiency of TERA/VCP/p97 contributes to impaired DNA repair in multiple polyglutamine diseases. *Nat. Commun.* **4**, 1816.
- Gabel, H. W., Kinde, B., Stroud, H., Gilbert, C. S., Harmin, D. A., Kastan, N. R., Hemberg, M., Ebert, D. H. and Greenberg, M. E. (2015). Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature* **522**, 89-93.
- Gacy, A. M., Goellner, G., Juranić, N., Macura, S. and McMurray, C. T. (1995). Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell* **81**, 533-540.
- Gao, R., Matsuura, T., Coolbaugh, M., Zühlke, C., Nakamura, K., Rasmussen, A., Siciliano, M. J., Ashizawa, T. and Lin, X. (2008). Instability of expanded CAG/CAA repeats in spinocerebellar ataxia type 17. *Eur. J. Hum. Genet.* **16**, 215-222.
- Gao, R., Liu, Y., Silva-Fernandes, A., Fang, X., Paulucci-Holthauzen, A., Chatterjee, A., Zhang, H. L., Matsuura, T., Choudhary, S., Ashizawa, T. et al. (2015). Inactivation of PNKP by mutant ATXN3 triggers apoptosis by activating the DNA damage-response pathway in SCA3. *PLoS Genet.* **11**, e1004834.
- GeM-HD Consortium (2015). Identification of genetic factors that modify clinical onset of huntington's disease. *Cell* **162**, 516-526.
- Gerber, H. P., Seipel, K., Georgiev, O., Höfferer, M., Hug, M., Rusconi, S. and Schaffner, W. (1994). Transcriptional activation modulated by homopolymeric glutamine and proline stretches. *Science* **263**, 808-811.
- Giuliano, P., de Cristofaro, T., Affaitati, A., Pizzulo, G. M., Feliciello, A., Crisuolo, C., De Michele, G., Filla, A., Avvedimento, E. V. and Varrone, S. (2003). DNA damage induced by polyglutamine-expanded proteins. *Hum. Mol. Genet.* **12**, 2301-2309.
- Goellner, G. M., Tester, D., Thibodeau, S., Almqvist, E., Goldberg, Y. P., Hayden, M. R. and McMurray, C. T. (1997). Different mechanisms underlie DNA instability in Huntington disease and colorectal cancer. *Am. J. Hum. Genet.* **60**, 879-890.
- Goldberg, Y. P., McMurray, C. T., Zeisler, J., Almqvist, E., Silence, D., Richards, F., Gacy, A. M., Buchanan, J., Telenius, H. and Hayden, M. R. (1995). Increased instability of intermediate alleles in families with sporadic Huntington disease compared to similar sized intermediate alleles in the general population. *Hum. Mol. Genet.* **4**, 1911-1918.
- Gonitel, R., Moffitt, H., Sathasivam, K., Woodman, B., Detloff, P. J., Faull, R. L. M. and Bates, G. P. (2008). DNA instability in postmitotic neurons. *Proc. Natl. Acad. Sci. USA* **105**, 3467-3472.
- Goula, A.-V., Berquist, B. R., Wilson, D. M., Wheeler, V. C., Trottier, Y. and Merienne, K. (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.
- Gövert, F. and Schneider, S. A. (2013). Huntington's disease and Huntington's disease-like syndromes. *Curr. Opin. Neurol.* **26**, 420-427.
- Gray, M., Shirasaki, D. I., Cepeda, C., Andre, V. M., Wilburn, B., Lu, X.-H., Tao, J., Yamazaki, I., Li, S.-H., Sun, Y. E. et al. (2008). Full-length human mutant huntingtin with a stable polyglutamine repeat can elicit progressive and selective neuropathogenesis in BACHD mice. *J. Neurosci.* **28**, 6182-6195.
- Gymrek, M., Willems, T., Guilmatre, A., Zeng, H., Markus, B., Georgiev, S., Daly, M. J., Price, A. L., Pritchard, J. K., Sharp, A. J. et al. (2015). Abundant contribution of short tandem repeats to gene expression variation in humans. *Nat. Genet.* **48**, 22-29.
- Hall, A. C., Ostrowski, L. A., Pietrobon, V. and Mekhail, K. (2017). Repetitive DNA loci and their modulation by the non-canonical nucleic acid structures R-loops and G-quadruplexes. *Nucleus* **8**, 162-181.
- Hashida, H., Goto, J., Kurisaki, H., Mizusawa, H. and Kanazawa, I. (1997). Brain regional differences in the expansion of a CAG repeat in the spinocerebellar ataxias: dentatorubral-pallidolysian atrophy, machado-joseph disease, and spinocerebellar ataxia type 1. *Ann. Neurol.* **41**, 505-511.
- Hensman-Moss, D. J., Pardiñas, A. F., Langbehn, D., Lo, K., Leavitt, B. R., Roos, R., Durr, A., Mead, S., TRACK-HD investigators; REGISTRY investigators, Holmans, P. et al. (2017). Identification of genetic variants associated with Huntington's disease progression: a genome-wide association study. *Lancet Neurol.* **16**, 701-711.
- Hoch, N. C., Hanzlikova, H., Rulten, S. L., Tétreault, M., Komulainen, E., Ju, L., Hornyak, P., Zeng, Z., Gittens, W., Rey, S. A. et al. (2016). XRCC1 mutation is associated with PARP1 hyperactivation and cerebellar ataxia. *Nature* **541**, 87-91.
- Hoelijmakers, J. H. J. (2009). DNA damage, aging, and cancer. *N. Engl. J. Med.* **361**, 1475-1485.
- Holmans, P. A., Massey, T. H. and Jones, L. (2017). Genetic modifiers of Mendelian disease: Huntington's disease and the trinucleotide repeat disorders. *Hum. Mol. Genet.* **24**, 55-60.
- Hubert, L., Lin, Y., Dion, V. and Wilson, J. H. (2011). Xpa deficiency reduces CAG trinucleotide repeat instability in neuronal tissues in a mouse model of SCA1. *Hum. Mol. Genet.* **20**, 4822-4830.
- Iyama, T. and Wilson, D. M. (2013). DNA repair mechanisms in dividing and non-dividing cells. *DNA Repair (Amst)* **12**, 620-636.
- Jackson, S. P. and Bartek, J. (2009). The DNA-damage response in human biology and disease. *Nature* **461**, 1071-1078.
- Jaworski, A., Rosche, W. A., Gellibolian, R., Kang, S., Shimizu, M., Bowater, R. P., Sinden, R. R. and Wells, R. D. (1995). Mismatch repair in Escherichia coli enhances instability of (CTG)<sub>n</sub> triplet repeats from human hereditary diseases. *Proc. Natl. Acad. Sci. USA* **92**, 11019-11023.
- Ji, J., Sundquist, K. and Sundquist, J. (2012). Cancer incidence in patients with polyglutamine diseases: a population-based study in Sweden. *Lancet. Oncol.* **13**, 642-648.
- Ju, B.-G., Lunyak, V. V., Perissi, V., Garcia-Bassets, I., Rose, D. W., Glass, C. K. and Rosenfeld, M. G. (2006). A topoisomerase II-mediated dsDNA break required for regulated transcription. *Science (80-)* **312**, 1798-1802.
- Jung, J. and Bonini, N. (2007). CREB-binding protein modulates repeat instability in a drosophila model for PolyQ disease. *Science (80-)* **315**, 1857-1859.
- Jung, Y., Mikata, Y. and Lippard, S. J. (2001). Kinetic studies of the TATA-binding protein interaction with cisplatin-modified DNA. *J. Biol. Chem.* **276**, 43589-43596.
- Kahle, J. J., Gulbahce, N., Shaw, C. A., Lim, J., Hill, D. E., Barabási, A.-L. and Zoghbi, H. Y. (2011). Comparison of an expanded ataxia interactome with patient medical records reveals a relationship between macular degeneration and ataxia. *Hum. Mol. Genet.* **20**, 510-527.
- Keeney, S., Lange, J. and Mohibullah, N. (2014). Self-organization of meiotic recombination initiation: general principles and molecular pathways. *Annu. Rev. Genet.* **48**, 187-214.
- Kennedy, L. (2003). Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. *Hum. Mol. Genet.* **12**, 3359-3367.
- Khan, N., Kolimi, N. and Rathinavelan, T. (2015). Twisting right to left: A...A mismatch in a CAG trinucleotide repeat overexpansion provokes left-handed Z-DNA conformation. *PLoS Comput. Biol.* **11**, e1004162.
- Kim, J. C., Harris, S. T., Dinter, T., Shah, K. A. and Mirkin, S. M. (2016). The role of break-induced replication in large-scale expansions of (CAG)<sub>n</sub>/(CTG)<sub>n</sub> repeats. *Nat. Struct. Mol. Biol.* **24**, 55-60.
- King, I. F., Yandava, C. N., Mabb, A. M., Hsiao, J. S., Huang, H.-S., Pearson, B. L., Calabrese, J. M., Starmer, J., Parker, J. S., Magnuson, T. et al. (2013). Topoisomerases facilitate transcription of long genes linked to autism. *Nature* **501**, 58-62.
- Koide, R., Ikeuchi, T., Onodera, O., Tanaka, H., Igarashi, S., Endo, K., Takahashi, H., Kondo, R., Ishikawa, A., Hayashi, T. et al. (1994). Unstable expansion of CAG repeat in hereditary dentatorubral-pallidolysian atrophy (DRPLA). *Nat. Genet.* **6**, 9-13.
- Kovalenko, M., Dragileva, E., St Claire, J., Gillis, T., Guide, J. R., New, J., Dong, H., Kucherlapati, R., Kucherlapati, M. H., Ehrlich, M. E. 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.
- Kovtun, I. V. and McMurray, C. T. (2001). Trinucleotide expansion in haploid germ cells by gap repair. *Nat. Genet.* **27**, 407-411.
- Kovtun, I. V., Johnson, K. O. and McMurray, C. T. (2011). Cockayne Syndrome B protein antagonizes OGG1 in modulating CAG repeat length *in vivo*. *Aging (Albany, NY)* **3**, 509-514.
- Lee, D.-Y. and McMurray, C. T. (2014). Trinucleotide expansion in disease: why is there a length threshold? *Curr. Opin. Genet. Dev.* **26**, 131-140.

- Lee, J.-M., Pinto, R. M., Gillis, T., St Claire, J. C. and Wheeler, V. C. (2011). Quantification of age-dependent somatic CAG repeat instability in Hdh CAG knock-in mice reveals different expansion dynamics in striatum and liver. *PLoS ONE* **6**, e23647.
- Libby, R. T., Monckton, D. G., Fu, Y.-H., Martinez, R. A., McAbney, J. P., Lau, R., Einum, D. D., Nichol, K., Ware, C. B., Ptacek, L. J. et al. (2003). Genomic context drives SCA7 CAG repeat instability, while expressed SCA7 cDNAs are intergenerationally and somatically stable in transgenic mice. *Hum. Mol. Genet.* **12**, 41-50.
- Lim, J., Hao, T., Shaw, C., Patel, A. J., Szabó, G., Rual, J.-F., Fisk, C. J., Li, N., Smolyar, A., Hill, D. E. et al. (2006). A protein-protein interaction network for human inherited ataxias and disorders of purkinje cell degeneration. *Cell* **125**, 801-814.
- Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature* **362**, 709-715.
- Litchfield, K., Shipley, J. and Turnbull, C. (2015). Common variants identified in genome-wide association studies of testicular germ cell tumour: an update, biological insights and clinical application. *Andrology* **3**, 34-46.
- Liu, J.-P. and Zeitlin, S. O. (2017). Is huntingtin dispensable in the adult brain? *J. Huntingtons. Dis.* **6**, 1-17.
- Lu, H.-C., Tan, Q., Rousseaux, M. W. C., Wang, W., Kim, J.-Y., Richman, R., Wan, Y.-W., Yeh, S.-Y., Patel, J. M., Liu, X. et al. (2017). Disruption of the ATXN1-CIC complex causes a spectrum of neurobehavioral phenotypes in mice and humans. *Nat. Genet.* **49**, 527-536.
- Madabhushi, R., Pan, L. and Tsai, L.-H. (2014). DNA damage and its links to neurodegeneration. *Neuron* **83**, 266-282.
- Madabhushi, R., Gao, F., Pfenning, A. R., Pan, L., Yamakawa, S., Seo, J., Rueda, R., Phan, T. X., Yamakawa, H., Pao, P.-C. et al. (2015). Activity-induced DNA breaks govern the expression of neuronal early-response genes. *Cell* **161**, 1592-1605.
- Maiuri, T., Mocle, A. J., Hung, C. L., Xia, J., van Roon-Mom, W. M. C. and Truant, R. (2016). Huntingtin is a scaffolding protein in the ATM oxidative DNA damage response complex. *Hum. Mol. Genet.* **26**, 395-406.
- Mangiarini, L., Sathasivam, K., Mahal, A., Mott, R., Seller, M. and Bates, G. P. (1997). Instability of highly expanded CAG repeats in mice transgenic for the Huntington's disease mutation. *Nat. Genet.* **15**, 197-200.
- Martins, S., Pearson, C. E., Coutinho, P., Provost, S., Amorim, A., Dubé, M.-P., Sequeiros, J. and Rouleau, G. A. (2014). Modifiers of (CAG)<sub>n</sub> instability in Machado-Joseph disease (MJD/SCA3) transmissions: an association study with DNA replication, repair and recombination genes. *Hum. Genet.* **133**, 1311-1318.
- McKinnon, P. J. (2013). Maintaining genome stability in the nervous system. *Nat. Neurosci.* **16**, 1523-1529.
- McKinnon, P. J. (2016). Topoisomerases and the regulation of neural function. *Nat. Rev. Neurosci.* **17**, 673-679.
- McMurray, C. T. (2010). Mechanisms of trinucleotide repeat instability during human development. *Nat. Rev. Genet.* **11**, 786-799.
- Menon, R. P., Nethisinghe, S., Faggiano, S., Vannocci, T., Rezaei, H., Pemble, S., Sweeney, M. G., Wood, N. W., Davis, M. B., Pastore, A. et al. (2013). The role of interruptions in polyQ in the pathology of SCA1. *PLoS Genet.* **9**, e1003648.
- Mirkin, S. M. (2007). Expandable DNA repeats and human disease. *Nature* **447**, 932-940.
- Møllersen, L., Rowe, A. D., Larsen, E., Rognes, T. and Klungland, A. (2010). Continuous and periodic expansion of CAG repeats in huntington's disease R6/1 mice. *PLoS Genet.* **6**, e1001242.
- Moshell, A. N., Tarone, R. E., Barrett, S. F. and Robbins, J. H. (1980). Radiosensitivity in Huntington's disease: implications for pathogenesis and presymptomatic diagnosis. *Lancet* **1**, 9-11.
- Nagafuchi, S., Yanagisawa, H., Sato, K., Shirayama, T., Ohsaki, E., Bundo, M., Takeda, T., Tadokoro, K., Kondo, I., Murayama, N. et al. (1994). Dentatorubral and pallidolysian atrophy expansion of an unstable CAG trinucleotide on chromosome 12p. *Nat. Genet.* **6**, 14-18.
- Nakatani, R., Nakamori, M., Fujimura, H., Mochizuki, H. and Takahashi, M. P. (2015). Large expansion of CTG-CAG repeats is exacerbated by MutSβ in human cells. *Sci. Rep.* **5**, 11020.
- Neil, A. J., Kim, J. C. and Mirkin, S. M. (2017). Precarious maintenance of simple DNA repeats in eukaryotes. *BioEssays* **39**, 1700077.
- Nelson, M. R., Tipney, H., Painter, J. L., Shen, J., Nicoletti, P., Shen, Y., Floratos, A., Sham, P. C., Li, M. J., Wang, J. et al. (2015). The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47**, 856-860.
- O'Brien, P. J. (2006). Catalytic promiscuity and the divergent evolution of DNA repair enzymes. *Chem. Rev.* **106**, 720-752.
- Ohno, M., Miura, T., Furuichi, M., Tominaga, Y., Tsuchimoto, D., Sakumi, K. and Nakabeppu, Y. (2006). A genome-wide distribution of 8-oxoguanine correlates with the preferred regions for recombination and single nucleotide polymorphism in the human genome. *Genome Res.* **16**, 567-575.
- Orr, H. T. (2012). Cell biology of spinocerebellar ataxia. *J. Cell Biol.* **197**, 167-177.
- Orr, H. T. and Zoghbi, H. Y. (2007). Trinucleotide repeat disorders. *Annu. Rev. Neurosci.* **30**, 575-621.
- Orr, H. T., Chung, M., Banfi, S., Kwiatkowski, T. J., Servadio, A., Beaudet, A. L., McCall, A. E., Duvick, L. A., Ranum, L. P. W. and Zoghbi, H. Y. (1993). Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nat. Genet.* **4**, 221-226.
- Ostrowski, L., Hall, A. and Mekhail, K. (2017). Ataxin-2: from RNA control to human health and disease. *Genes (Basel)* **8**, 157.
- Panigrahi, G. B., Lau, R., Montgomery, S. E., Leonard, M. R. and Pearson, C. E. (2005). Slipped (CTG)<sub>n</sub>(CAG) repeats can be correctly repaired, escape repair or undergo error-prone repair. *Nat. Struct. Mol. Biol.* **12**, 654-662.
- Panigrahi, G. B., Sleam, M. M., Simard, J. P., Gileadi, O. and Pearson, C. E. (2010). Isolated short CTG/CAG DNA slip-outs are repaired efficiently by hMutSbeta, but clustered slip-outs are poorly repaired. *Proc. Natl. Acad. Sci. USA* **107**, 12593-12598.
- Pearl, L. H., Schierz, A. C., Ward, S. E., Al-Lazikani, B. and Pearl, F. M. G. (2015). Therapeutic opportunities within the DNA damage response. *Nat. Rev. Cancer* **15**, 166-180.
- Pearson, C. E., Ewel, A., Acharya, S., Fishel, R. A. and Sinden, R. R. (1997). Human MSH2 binds to trinucleotide repeat DNA structures associated with neurodegenerative diseases. *Hum. Mol. Genet.* **6**, 1117-1123.
- Pearson, C. E., Eichler, E. E., Lorenzetti, D., Kramer, S. F., Zoghbi, H. Y., Nelson, D. L. and Sinden, R. R. (1998). Interruptions in the triplet repeats of SCA1 and FRAXA reduce the propensity and complexity of slipped strand DNA (S-DNA) formation. *Biochemistry* **37**, 2701-2708.
- Pinto, R. M., Dragileva, E., Kirby, A., Lloret, A., Lopez, E., St Claire, J., Panigrahi, G. B., Hou, C., Holloway, K., Gillis, T. 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.
- Plenge, R. M., Scolnick, E. M. and Altshuler, D. (2013). Validating therapeutic targets through human genetics. *Nat. Rev. Drug Discov.* **12**, 581-594.
- Ripke, S., Neale, B. M., Corvin, A., Walters, J. T. R., Farh, K.-H., Holmans, P. A., Lee, P., Bulik-Sullivan, B., Collier, D. A., Huang, H. et al. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-427.
- Robison, S. H. and Bradley, W. G. (1984). DNA damage and chronic neuronal degenerations. *J. Neurol. Sci.* **64**, 11-20.
- Rodriguez, G. P., Romanova, N. V., Bao, G., Rouf, N. C., Kow, Y. W. and Crouse, G. F. (2012). Mismatch repair-dependent mutagenesis in nondividing cells. *Proc. Natl. Acad. Sci. USA* **109**, 6153-6158.
- Ruano, L., Melo, C., Silva, M. C. and Coutinho, P. (2014). The global epidemiology of hereditary ataxia and spastic paraplegia: a systematic review of prevalence studies. *Neuroepidemiology* **42**, 174-183.
- Rué, L., Bañez-Coronel, M., Creus-Muncunill, J., Giral, A., Alcalá-Vida, R., Mentxaka, G., Kagerbauer, B., Zomeño-Abellán, M. T., Aranda, Z., Venturi, V. et al. (2016). Targeting CAG repeat RNAs reduces Huntington's disease phenotype independently of huntingtin levels. *J. Clin. Invest.* **126**, 4319-4330.
- Rulten, S. L. and Caldecott, K. W. (2013). DNA strand break repair and neurodegeneration. *DNA Repair (Amst)* **12**, 558-567.
- Sánchez, I., Balagué, E. and Matilla-Dueñas, A. (2016). Ataxin-1 regulates the cerebellar bioenergetics proteome through the GSK3β-mTOR pathway which is altered in Spinocerebellar ataxia type 1 (SCA1). *Hum. Mol. Genet.* **25**, 4021-4040.
- Saudou, F. and Humbert, S. (2016). The biology of huntingtin. *Neuron* **89**, 910-926.
- Scarborough, P. M., Weber, R. P., Iversen, E. S., Brhane, Y., Amos, C. I., Kraft, P., Hung, R. J., Sellers, T. A., Witte, J. S., Pharoah, P. et al. (2016). A cross-cancer genetic association analysis of the DNA repair and DNA damage signaling pathways for lung, ovary, prostate, breast, and colorectal cancer. *Cancer Epidemiol. Biomarkers Prev.* **25**, 193-200.
- Scudiero, D. A., Meyer, S. A., Clatterbuck, B. E., Tarone, R. E. and Robbins, J. H. (1981). Hypersensitivity to N-methyl-N'-nitro-N-nitrosoguanidine in fibroblasts from patients with Huntington disease, familial dysautonomia, and other primary neuronal degenerations. *Proc. Natl. Acad. Sci. USA* **78**, 6451-6455.
- Sequeiros, J., Seneca, S. and Martindale, J. (2010). Consensus and controversies in best practices for molecular genetic testing of spinocerebellar ataxias. *Eur. J. Hum. Genet.* **18**, 1188-1195.
- Shadyab, A. H. and LaCroix, A. Z. (2015). Genetic factors associated with longevity: a review of recent findings. *Ageing Res. Rev.* **19**, 1-7.
- Shelbourne, P. F., Keller-McGandy, C., Bi, W. L., Yoon, S. R., Dubeau, L., Veitch, N. J., Vonsattel, J. P., Wexler, N. S., Norman, A. and Augood, S. J. (2007). Triplet repeat mutation length gains correlate with cell-type specific vulnerability in Huntington disease brain. *Hum. Mol. Genet.* **16**, 1133-1142.
- Sleam, M. M., Panigrahi, G. B., Ranum, L. P. and Pearson, C. E. (2008). Mutagenic roles of DNA 'repair' proteins in antibody diversity and disease-associated trinucleotide repeat instability. *DNA Repair (Amst)* **7**, 1135-1154.
- Sobczak, K. and Krzyzosiak, W. J. (2004). Imperfect CAG repeats form diverse structures in SCA1 transcripts. *J. Biol. Chem.* **279**, 41563-41572.
- Song, J., Bergen, S. E., Di Florio, A., Karlsson, R., Charney, A., Ruderfer, D. M., Stahl, E. A., Members of the International Cohort Collection for Bipolar Disorder (ICCBD), Chambert, K. D., Moran, J. L. et al. (2016). Genome-wide association study identifies SESTD1 as a novel risk gene for lithium-responsive bipolar disorder. *Mol. Psychiatry* **21**, 1290-1297.
- Spada, A. R. L., Wilson, E. M., Lubahn, D. B., Harding, A. E. and Fischbeck, K. H. (1991). Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* **352**, 77-79.

- Stevens, J. R., Lahue, E. E., Li, G.-M. and Lahue, R. S. (2013). Trinucleotide repeat expansions catalyzed by human cell-free extracts. *Cell Res.* **23**, 565-572.
- Suberbielle, E., Sanchez, P. E., Kravitz, A. V., Wang, X., Ho, K., Eilertson, K., Devidze, N., Kreitzer, A. C. and Mucke, L. (2013). Physiologic brain activity causes DNA double-strand breaks in neurons, with exacerbation by amyloid- $\beta$ . *Nat. Neurosci.* **16**, 613-621.
- Sun, Y.-M., Lu, C. and Wu, Z.-Y. (2016). Spinocerebellar ataxia: relationship between phenotype and genotype - a review. *Clin. Genet.* **90**, 305-314.
- Swami, M., Hendricks, A. E., Gillis, T., Massood, T., Mysore, J., Myers, R. H. and Wheeler, V. C. (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**, 3039-3047.
- Synofzik, M. and Schüle, R. (2017). Overcoming the divide between ataxias and spastic paraplegias: Shared phenotypes, genes, and pathways. *Mov. Disord.* **32**, 332-345.
- Tanaka, F., Sobue, G., Doyu, M., Ito, Y., Yamamoto, M., Shimada, N., Yamamoto, K., Riku, S., Hshizume, Y. and Mitsuma, T. (1996). Differential pattern in tissue-specific somatic mosaicism of expanded CAG trinucleotide repeat in dentatorubral-pallidolusian atrophy, Machado-Joseph disease, and X-linked recessive spinal and bulbar muscular atrophy. *J. Neurol. Sci.* **135**, 43-50.
- Taniguchi, J. B., Kondo, K., Fujita, K., Chen, X., Homma, H., Sudo, T., Mao, Y., Watake, K., Tanaka, T., Tagawa, K. et al. (2016). RPA1 ameliorates symptoms of mutant ataxin-1 knock-in mice and enhances DNA damage repair. *Hum. Mol. Genet.* **25**, 4432-4447.
- Telenius, H., Almqvist, E., Kremer, B., Spence, N., Squitieri, F., Nichol, K., Grandell, U., Starr, E., Benjamin, C. and Castaldo, I. (1995). Somatic mosaicism in sperm is associated with intergenerational (CAG)<sub>n</sub> changes in Huntington disease. *Hum. Mol. Genet.* **4**, 189-195.
- The Huntington's Disease Collaborative Research Group (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group. *Cell* **72**, 971-983.
- Usdin, K., House, N. C. M. and Freudenreich, C. H. (2015). Repeat instability during DNA repair: insights from model systems. *Crit. Rev. Biochem. Mol. Biol.* **50**, 142-167.
- Volle, C. B., Jarem, D. A. and Delaney, S. (2012). Trinucleotide Repeat DNA Alters Structure To Minimize the Thermodynamic Impact of 8-Oxo-7,8-dihydroguanine. *Biochemistry* **51**, 52-62.
- Wang, L. and Dent, S. Y. R. (2014). Functions of SAGA in development and disease. *Epigenomics* **6**, 329-339.
- Wang, G., Sawai, N., Kotliarova, S., Kanazawa, I. and Nukina, N. (2000). Ataxin-3, the MJD1 gene product, interacts with the two human homologs of yeast DNA repair protein RAD23, HHR23A and HHR23B. *Hum. Mol. Genet.* **9**, 1795-1803.
- Ward, J. M. and La Spada, A. R. (2015). Ataxin-3, DNA damage repair, and SCA3 cerebellar degeneration: on the path to parsimony? *PLoS Genet.* **11**, e1004937.
- Wardle, M., Morris, H. R. and Robertson, N. P. (2009). Clinical and genetic characteristics of non-Asian dentatorubral-pallidolusian atrophy: a systematic review. *Mov. Disord.* **24**, 1636-1640.
- Weissglas-Volkov, D., Aguilar-Salinas, C. A., Nikkola, E., Deere, K. A., Cruz-Bautista, I., Arellano-Campos, O., Muñoz-Hernandez, L. L., Gomez-Munguia, L., Ordoñez-Sánchez, M. L., Reddy, P. M. V. L. et al. (2013). Genomic study in Mexicans identifies a new locus for triglycerides and refines European lipid loci. *J. Med. Genet.* **50**, 298-308.
- Wexler, N. S., Lorimer, J., Porter, J., Gomez, F., Moskowitz, C., Shackell, E., Marder, K., Penchaszadeh, G., Roberts, S. A., Gayan, J. et al. (2004). Venezuelan kindreds reveal that genetic and environmental factors modulate Huntington's disease age of onset. *Proc. Natl. Acad. Sci. USA* **101**, 3498-3503.
- Wheeler, V. C., Lebel, L.-A., Vrbanac, V., Teed, A., te Riele, H. and MacDonald, M. E. (2003). Mismatch repair gene Msh2 modifies the timing of early disease in HdhQ111 striatum. *Hum. Mol. Genet.* **12**, 273-281.
- Williams, G. M. and Surtees, J. A. (2015). MSH3 promotes dynamic behavior of trinucleotide repeat tracts in vivo. *Genetics* **200**, 737-754.
- Xiao, H., Yu, Z., Wu, Y., Nan, J., Merry, D. E., Sekiguchi, J. M., Ferguson, D. O., Lieberman, A. P. and Dressler, G. R. (2012). A polyglutamine expansion disease protein sequesters PTIP to attenuate DNA repair and increase genomic instability. *Hum. Mol. Genet.* **21**, 4225-4236.
- Yu, Z., Zhu, Y., Chen-Plotkin, A. S., Clay-Falcone, D., McCluskey, L., Elman, L., Kalb, R. G., Trojanowski, J. Q., Lee, V. M.-Y., Van Deerlin, V. M. et al. (2011). PolyQ repeat expansions in ATXN2 associated with ALS Are CAA interrupted repeats. *PLoS ONE* **6**, e17951.
- Zeitlin, S., Dragatsis, I. and Levine, M. S. (2000). Inactivation of Hdh in the brain and testis results in progressive neurodegeneration and sterility in mice. *Nat. Genet.* **26**, 300-306.
- Zylka, M. J., Simon, J. M. and Philpot, B. D. (2015). Gene length matters in neurons. *Neuron* **86**, 353-355.