Title: Tissue specificity in DNA repair: lessons from trinucleotide repeat instability.
Authors: Dion V
Journal: Trends in genetics
Year: 2014
Volume: 30
Issue: 6
Pages: 220-9
DOI: 10.1016/j.tig.2014.04.005

Tissue specificity in DNA repair:
Lessons from trinucleotide repeat instability

Vincent Dion
University of Lausanne
Center for Integrative Genomics
Bâtiment Génopode
1015 Lausanne
Switzerland
vincent.dion@unil.ch

Keywords: DNA repair, genome stability, trinucleotide repeat instability, Base excision repair, single-strand break repair, nucleotide excision repair, tissue-specific DNA repair.

Abstract

DNA must constantly be repaired to maintain genome stability. Although it is clear that DNA repair reactions depend on cell types and developmental stages, we know surprisingly little about the mechanisms that underlie tissue-dependence. This is due, in part, to the lack of adequate study systems. This review discusses recent progress towards understanding the mechanism leading to varying rates of instability at expanded trinucleotide repeats (TNRs) in different tissues. Although they are not DNA lesions, TNRs are hotspots for genome instability because normal DNA repair activities cause changes in repeat length. The rates of expansions and contractions are readily detectable and depend on cell identity, making TNR instability a particularly convenient model system. A better understanding of this type of genome instability will provide a foundation for studying tissue-specific DNA repair more generally, which has implications in cancer and other diseases caused by mutations in the genome's caretakers.

Handling DNA damage in a tissue-dependent manner

DNA repair exhibits tissue-specific differences that have received surprisingly little attention despite our thorough characterization of the pathways that handle the 100 000 lesions made to DNA per cell per day [1] (see Figure 1 for details of the repair pathways discussed in this review). Yet this aspect of DNA repair has profound implications on how different cells maintain genome stability and on how we treat specific cancers [2, 3]. Most of what we know about DNA repair mechanisms comes from experiments performed in single-cell models, such as bacteria, yeast, and cultured mammalian cells, or from biochemical assays that use cell extracts or purified proteins. Although these model systems are easily manipulated and are ideal to study mutagenesis, they are ill suited for uncovering mechanisms governing tissue-specific differences in DNA repair. Here I discuss what is known about the cell-type specificity of DNA repair and focus on one particularly suitable system to study this phenomenon: the instability of expanded trinucleotide repeats (TNRs). TNRs are hotspots for genome instability under otherwise normal conditions and their very high mutation rates depend on a variety of DNA repair pathways, thereby providing a convenient site-specific system to study the dissimilarities in DNA repair between tissues.

Several lines of evidence suggest that DNA damage is handled differently between tissues. First, patients with mutations in key DNA repair genes often have tissue-specific phenotypes (Table 1) [3]. For instance, mutations in the mismatch repair (MMR) pathway, specifically those that inactivate MSH2, MSH6, MLH1, or PMS2, lead to colorectal cancer, whereas mutations in proteins involved in single-strand break repair (SSBR), tend to impair neurological functions, as is the case for aprataxin mutations that lead to ataxia with oculomotor apraxia [3]. These differences might arise because MMR is particularly important in dividing cells since mismatches occur at the replication fork. By contrast, neurons would be more sensitive to single-strand breaks and oxidative damage because they are postmitotic and have high metabolic rates [3].

Interestingly, sequencing cancer genomes has revealed that tumours show a wide range of mutational signatures that vary in part because of their tissue of origin [4]. Mouse knockouts provide further evidence that specific repair pathways are often crucial for the development and continued function of only a subset of cell types [3, 5].

Measuring the rates of mutations in mouse models carrying a reporter gene also indicate that DNA repair is tissue-dependent. For instance, there is a three-fold difference in the rate of spontaneous mutations within a LacI transgene between germ cells and the spleen [6]. Similarly, fibroblasts are 100 times more prone to mutations than embryonic stem cells as measured by selecting cells for the functional loss of the non-essential *APRT* gene [7]. These observations show that mutation rates are indeed tissue-specific.

Differentiating mammalian cell lines into neuron-like cells *in vitro* revealed an overall reduction in DNA repair capability, yet repair within transcribed domains becomes more efficient, probably via an ill-defined variant of the transcription-coupled repair (TCR) pathway, [8, 9]. The assays used in these studies have the advantage that the dose and type of DNA damage can be controlled and differentiation can be induced at will. Yet, they are limited because the exact location of DNA damage is unknown, a factor that influences DNA repair and genome instability [10].

Biochemical approaches have also been used to assess directly DNA repair differences in various tissues. For instance, a thorough study evaluated the repair of oxidative lesions via base excision repair (BER) in nuclear extracts from six mouse tissues using oligonucleotide substrates harbouring a single DNA lesion [11]. In this system, the excision of lesions by DNA glycosylases is highest in the testes, providing strong evidence that BER is indeed tissue-specific. Although these types of experiments exclude the effects of chromatin structure, which is both tissue-specific [12] and impacts the repair of DNA lesions [13], their use of a defined substrate allows for precise quantification of repair efficiencies.

Some studies have focused on the amount of DNA damage accumulating between mouse tissues. For example, the steady state level of several oxidized DNA bases is higher in the brain than in the liver and tail tissues, especially in older mice [14]. Such experiments may be interpreted as evidence that DNA repair differs between tissues or that the amounts of the DNA lesions incurred varies in different cell types.

Trinucleotide repeat instability provides an opportunity to study tissue-specific DNA repair

The experiments outlined above clearly indicate that DNA lesions have tissue-selective effects, but the underlying mechanisms are not well understood and have been difficult to tease apart. Microsatellites, defined as tandem repeats of up to 6 nucleotides, offer a more tractable system for investigating the tissue-specificity of DNA repair and are particularly important because they are often associated with neurological disorders [15]. Here I focus on CAG/CTG repeats because the mechanisms of their tissue-dependent instability have been better characterized and many mouse models are available that recapitulate most of the features of instability seen in human patients, albeit some differences are present depending on the exact mouse model used [16]. The instability comes in two forms: expansions and contractions, which include the addition or deletion of repeat units, respectively. TNRs become highly unstable once they reach beyond a threshold of about 35 units, above which they can efficiently adopt non-B structures, such as hairpins and slipped strands, that can be bound by DNA repair factors or bypassed by the DNA synthesis machinery, leading to contraction and expansion of the number of repeats at the locus (Fig. 2) [17-19]. Thus, any factors that promote secondary structure formation or that contribute to repair of the repeat tract are involved in TNR instability. The genetic factors known to be important for TNR instability are thus central players in nearly every repair pathway, including TCR, MMR, BER, SSBR, and, at least in lower eukaryotes, homologous recombination [17, 19]. DNA replication is also a major contributor of TNR instability, especially in lower eukaryotes [17, 19]. Thus, TNR instability represents a unique system to understand how the normal DNA repair machinery handles difficult sequences in the genome.

Repeat instability has several advantages for studying tissue-selective phenotypes in DNA repair. First, the site of the instability is defined, which allows the study of the contribution of cis-elements: chromatin-immunoprecipitation (ChIP) assays can be performed (e.g., [20-22]), the position of the nearest replication origins can be mapped (e.g., [23-26]), and the amount of mRNA expressed from the locus can be quantified, all of which can be tissue-dependent. Furthermore, the outcome of DNA repair at this locus, TNR instability, can be quantitatively measured (box 1). Importantly, the instability is tissue-specific, with some tissues having very high levels of instability (e.g., striatum and liver), while others, such as the heart and cerebellum, displaying relatively low instability. This tissue-specificity is surprisingly consistent between different mouse models that carry long CAG/CTG repeats at different loci in the genome (See Table 2 and 3). Together, these features make CAG/CTG instability in the mouse an especially useful *in vivo* model for elucidating the mechanisms guiding tissue-specific DNA repair.

Tissue-specific effects of genetic factors that affect TNR instability

Although some mouse knockouts, including NEIL1$^{-/-}$ [27], PMS2$^{-/-}$ [28], MSH2$^{-/-}$ [21, 29-36], MSH3$^{-/-}$ [35, 37, 38], MLH1$^{-/-}$ [39], and MLH3$^{-/-}$ [39], display changes in TNR instability in all tissues examined, many studies have identified genetic factors that impact TNR instability in either somatic or germ cells, but not in both. For instance, the DNA methyltransferase Dnmt1 prevents repeat expansion in the germline of both male and female SCA1 mice, but not in their somatic tissues [22]. DNA Ligase 1 (LIG1), which is involved in ligating nicks created during DNA replication, SSBR, and BER, is only required for CTG instability in the female germline of DM300 mice [40]. Knocking out one copy of FEN1, an endonuclease involved in processing DNA replication, SSBR, and BER intermediates, only impacted TNR instability in the male germline of the R6/1 mouse [41]. OGG1, a DNA glycosylase that excises 8-oxo-guanine and initiates BER, contributes to TNR instability in the somatic tissues of R6/1 mice, but has no measurable effect in the germline [14, 42]. These results highlight that TNR instability in somatic and germline cells involve different genetic players and may therefore occur via distinct mechanisms.

Differences in tissue specificity are also evident between somatic tissues. For example, different dynamics of expansions can be observed in the liver and striatum of the Hdh^{Q111} mouse [43]. A study found that SCA1 mice lacking XPA, a gene required for NER, had a stabilized TNR in the striatum, hippocampus, and cerebral cortex, but had no effect in the liver and kidneys [44]. These striking observations imply that the mechanism of instability is different even between somatic tissues.

Models for tissue-specificity of TNR instability

Every step leading to TNR instability could be tissue-specific, from the initiation of repair to the processing of intermediates. There is no consensus explaining the cause of the tissue-specificity, yet a number of hypotheses have been proposed [17, 19, 45]. They can be divided into four broad categories: 1) the frequency at which repair within the repeat tract is initiated; 2) the varying stoichiometry of DNA repair proteins; 3) the rate of replication and the location of nearby origins of replication; and 4) transcription and chromatin structure. Below I discuss these models individually, yet it is important to stress that they are by no means mutually exclusive. These models can also account for tissue-selectivity of DNA repair events in general and are not limited to TNR instability.

1) Frequency at which repair within the repeat tract is initiated

An important determinant of tissue-specific DNA repair is likely to be the type and amount of DNA lesions present in the different tissues. In the context of TNRs, this has been proposed to play an important role in BER-dependent instability. A specific type of oxidized base, 8-oxo-guanine (8-oxoG), accumulates in an age- and tissue-dependent manner in R6/1 mice [14] and expanded TNRs may be particularly prone to oxidative damage [46]. This accumulation was proposed to trigger the repair of the repeat tract, which would in turn increase the chances of aberrant repair and thus lead to instability. This model is supported by experiments with mice deficient for OGG1, which removes 8-oxoG from DNA. OGG1 deletion reduced instability in all somatic tissues, suggesting that the removal of 8-oxoG triggers instability [14]. Similarly,

deletion of NEIL1, a glycosylase that excises preferentially oxidized pyrimidines, but also 8-oxoG, reduced CAG/CTG instability in the R6/1 mice [27]. Together, these studies suggest that the repair of oxidized bases causes instability and predicts that the amount of damaged DNA determines instability rates in different tissues.

This model has been challenged, however, by the finding that the cerebellum accumulates high levels of abasic sites but has little instability in both the R6/1 and R6/2 mice [47]. The striatum, on the other hand, has fewer abasic sites but much more instability. In addition, nuclear extracts from the cerebellum excise 8-oxoG with higher efficiencies than other brain regions [48]. This goes counter to the model outlined above and suggests that tissue-specific differences in TNR instability also depend on other factors. It is possible that the frequency of mistakes made in the process of repairing an oxidized base near or within the CAG/CTG repeat is different between the two organs. For example, more 8-oxoG is encountered in the cerebellum, but its repair could be mostly error-free. In the striatum, there may be less oxidative damage, but the frequency of misrepair may be very high, leading to high levels of instability. Put differently, the number of lesions within the repeat tract, the frequency with which these lesions are repaired, and the frequency of misrepair are all likely to contribute to instability.

Virtually all models of repeat instability center around the ability of TNRs to form secondary structures that are recognized by DNA repair proteins [17, 19]. Analogous to DNA lesions within or near an expanded TNR, secondary structure formation could initiate repair and lead to instability. Larger amounts of slipped-strands could be immunoprecipitated using an antibody against three-way DNA junctions from the heart compared to the cerebellum of a patient with an expanded TNR at the *DMPK* locus [49]. This raises the possibility that structure formation is tissue-dependent and/or that the number of repeats was higher in the heart (4100-6000 CTGs) than in the cerebellum (1100-1500), creating proportionally different amounts of secondary structures.

Together, these studies suggest that the rate at which repair is initiated within the repeat tract, whether it is due to DNA damage or to the formation of secondary structure could account in part for the tissue-specificity. Going forward, we will need more quantitative data points in a larger set of tissues to evaluate this model. Such experiments would also shed light on how DNA repair of a variety of lesions, not only TNRs, occurs in different tissues.

2) Variations in the stoichiometry of DNA repair proteins across tissues

One of the predominant models that aims to explain tissue-specificity in TNR instability stipulates that the levels of DNA repair factors directly impact repeat instability. Changing the relative levels of DNA repair factors may potentially affect DNA repair rates.

The mismatch repair proteins MSH2, MSH3, MLH1, MLH3, and PMS2 have all been shown to modulate TNR instability in mouse models with CAG, CTG, GAA, and CGG expansions [21, 29-39, 50-53]. In addition, MSH6 is involved in GAA repeat instability but appears to have little effect on CAG/CTG instability [35, 38, 51-53]. These results prompted a systematic investigation of the levels of MSH2, MSH3 and MSH6 in several murine tissues [54], but no obvious correlation was found between instability phenotypes and protein levels. A recent study with the R6/1 mouse proposed that there is a “sweet spot” for instability where a moderate level of MSH2 and MSH3 leads to maximal instability whereas both very high and very low amounts prevent instability [55]. It is unclear how this would work mechanistically. These results highlight that simple correlations will probably not hold true and that complex mechanisms are at play. For instance, in addition to the levels of these proteins, their subcellular location and posttranslational modifications could alter their function in a tissue-specific manner. For example, MSH2 is acetylated [56] and this acetylation has been speculated to alter MSH2-dependent TNR instability in human cells [57].

The level of BER pathway components may play a role in defining tissue-specificity of instability. For instance, there are low levels of FEN1 and HMGB1, two factors that favour the long-patch

(LP) BER pathway, in the striatum, whereas there are high levels in the cerebellum where TNR instability is very high and very low, respectively [47]. This simple inverse correlation was substantiated by *in vitro* assays in which varying the amount of BER proteins to mimic the striatal or cerebellar situations modulated the processing of a CAG/CTG repeat by LP-BER and thereby the instability [58]. Together these experiments strongly suggest that the amount of BER proteins can account for the tissue-specificity of instability. One should be careful, however, as the levels of BER factors have not been determined for a large variety of tissues and a correlation between two tissues may not hold true in others. Nonetheless, this is an exciting possibility and cause-effect experiments are worth attempting.

Transcription-coupled repair has been implicated in CAG/CTG instability in bacteria [59, 60], human cells [61-63], and *Drosophila* [64]. Follow up experiments in mouse models carrying expanded CAG/CTG repeats showed that both XPA and CSB, but not XPC – a component of global genome NER pathway – had instability phenotypes [42, 44, 53]. $XPA^{-/-}$ mice showed a stabilization of the repeat tract only in neuronal tissues, arguing strongly that XPA has tissue-specific roles in the process. Curiously the upstream factor, CSB, does not appear to have specific roles in neuronal instability as would have been expected of a factor working upstream of XPA. This seeming incongruity may be because of the mouse model used (R6/1 versus SCA1), or the low number of mice analysed when conducting the CSB experiments, and/or that CSB has an XPA-independent role [44, 65].

The levels of NER proteins have not, to my knowledge, been investigated thoroughly in different mouse tissues [8, 9]. However, differentiation systems in cultured human cells showed that although the capacity of neuronal-like cells to repair UV-induced damage was decreased overall, TC-NER was buttressed both on the transcribed and non-transcribed strands [8, 9]. This unusual behaviour in differentiated cells may explain why XPA is critically important for TNR instability in these non-dividing tissues.

A recent genome-wide study found no correlation between the mRNA levels of DNA repair genes and repeat instability [66]. Instead, cell cycle transcripts and genes involved in neurotransmission correlated with repeat instability. This is intriguing and at the moment it is unclear whether the products of these genes have functions that would affect the stability of TNRs. Although such systems-biology approaches will be key to understanding the tissue-specificity of TNR instability, one clear caveat is that mRNA levels do not always correlate with protein levels. This is true both globally [67] and for specific MMR and BER components [47, 54].

3) Replication as a source of tissue-specific instability

Pioneering studies in bacteria and yeast identified replication as a major source of instability [68]. In mammalian cells, however, it appears that although replication can be linked to repeat instability, the changes tend to be relatively small [69-71]. Moreover, post-mitotic neurons display instability, suggesting the existence of replication-independent mechanisms. It was pointed out early that replication rates do not correlate with TNR instability in various organs [72]. By contrast, a recent study in the R6/2 mice shows that mRNA and protein levels of the DNA replication genes PCNA, FEN1, RPA1, and LIG1 are higher in the cerebellum than in the striatum, which have modest and high rates of TNR instability, respectively. Using instability and expression data from many more tissues from the Hdh^{Q111} mice, however, no such correlation has been noted [66]. Nevertheless, replication could potentially account for instability in proliferating and developing tissues. Indeed, tissue-specific changes have been observed in the position of replication origins with respect to the TNR in tissues from the DM300 mouse [23] (Table 1). Even though these changes do not correlate well with instability rates between tissues, it was proposed that changes in the position of the origin of replication could explain, at least in part, the instability seen in testes, where a large fraction of the cells proliferate [23].

4) Transcription and chromatin structure in TNR instability

Chromatin structure varies from tissue to tissue and influences gene expression, which ultimately determines cell identity [12]. It is possible that chromatin state underpins the tissue-specific phenotypes associated with TNR instability by, for example, modulating gene expression of DNA repair genes in different tissues. It is also possible that the differences in chromatin structure at TNR between tissues differentially affect their instability [73].

Transcription has long been linked to genome instability and transcribed regions of the genome appear to be particularly susceptible to breakage and mutations [74]. The TCR branch of NER is dedicated to the removal of DNA lesions that block RNA polymerase II (RNAPII) [75]. Transcription through a CAG/CTG repeat was not considered to be a major determinant of tissue-specific instability because genes containing expanded TNRs are ubiquitously expressed and the steady-state levels of their mRNAs do not correlate with tissue-specific instability [72, 76, 77]. Importantly, transcription elongation or initiation, rather than steady-state levels of the mRNAs, may still show a correlation. Therefore a series of experiments in mammalian cells was designed to test the hypothesis that transcription itself promotes repeat instability. In these studies, an inducible promoter driving transcription through the CAG/CTG tract increased rates of large contractions by 15 fold [61]. In addition, *in vitro* experiments show RNAPII stalling at hairpins formed by TNRs, which is thought to initiate TCR and trigger instability [78]. Further experiments in mammalian cells and in the *Drosophila* germline confirmed that transcription is important for expansion as well as contraction of TNRs [63, 64].

Recently, a specific histone modification that marks transcription elongation, trimethylation of histone H3 on lysine 36 (H3K36me3), was found to correlate with TNR instability in the cerebellum and striatum of the R6/1 and R6/2 mice [79]. Moreover, the levels of RNAPII, as measured by ChIP, were also higher in the striatum than in the cerebellum, suggesting that there is a correlation between transcription elongation at the repeat locus and the levels of repeat instability in these two different mouse models. Interestingly, H3K36me3 is also known to affect DNA repair and recombination [80]. It is therefore tempting to speculate that this finding may have wider implications for genome instability.

That study is particularly interesting in light of several other experiments investigating the potential role of chromatin structure, including histone modifications, DNA methylation, and the presence of the boundary factor CTCF in repeat instability [22, 57, 64, 73, 79, 81, 82]. In mice, changes in DNA methylation near the expanded CAG/CTG repeat tract in the SCA1 knock-in mouse correlated with repeat instability in the germline, but not in somatic tissues [22]. In the SCA7-CTCF-I-wt mice, one out of 15 mice analyzed displayed high levels of TNR instability in its kidney, but had wild type levels in other tissues [82]. This mimicked the situation when the CTCF binding site in SCA7-CTCF-I-mut mice was mutated [82]. Upon further analysis, it was found that the DNA in this mouse's kidney was hypermethylated, which prevented CTCF binding *in vitro* [82]. These data suggest that changes in DNA methylation near the repeat tract have the potential to lead to tissue-specific differences in repeat instability. The contribution of chromatin structure to repeat instability remains understudied and may provide further clues as to how tissue-selectivity is achieved.

Concluding remarks

It is clear that the issue of tissue-specificity in DNA repair is complex and that simple correlations are unlikely to be found. More likely, several aspects of each of the models discussed above will be at play and may lead to synergistic interactions that regulate the observed instability.

The data accumulated thus far is largely correlational, yet they have provided a testable set of hypotheses. Going forward, I propose that two lines of research will help illuminate these issues. First, research in the field must move from correlative observations to cause-effect experiments. Granted this is more easily said than done. Generating mouse models where the amount of repair proteins can be modulated in a tissue-specific manner would be an informative start.

Second, most studies so far have focused on only a handful of proteins and histone modifications, which inevitably reduces the scope of the conclusions. More studies at the systems level (such as [66]) are needed. The use of systems biology approaches will provide insights into the pathways leading to repeat instability and the interactions between them. Such approaches, of course, are applicable to other types of genome instability, and the general problem of tissue-specific DNA repair events will also benefit from their results.

Studying the tissue-specificity of DNA repair will help our understanding of what goes astray in cancer, in the tissue-selectivity of several disorders caused by mutations in DNA repair genes [83], and in the neurological diseases caused by expanded repeats [15].

Acknowledgements

I thank Fisun Hamaratoglu, Gustavo A. Ruiz Buendia, Marie Sgandurra, and Andrew Seeber for critical reading of the manuscript. A Swiss National Science Foundation Professorship #PP00P3_144789 to V.D. provides the funds for this research.

Table 1: Genetic diseases caused by mutation in SSBR, MMR, BER, and NER repair genes and their tissue-specific effects.

Disease	OMIM number	Mutated genes	Pathway	Symptoms	Tissues affected	References
Ataxia with Oculomotor apraxia	208920	APTX	SSBR	Peripheral axonal neuropathy, oculomotor apraxia, hypoalbuminemia.	Cerebellum, neurons, muscles.	[84]
Cockayne syndrome	216400,133540 610651,278760 278780	CSA, CSB, XPB XPF, XPG	TC-NER	Developmental delay, dwarfism, photosensitivity, progressive pigmentory retinopathy, neurodegeration and mental retardation.	Skin, eyes, retina, cerebellum, cerebrum, neurons.	[85]
DNA ligase 1 deficiency	126391.0001	LIG1	BER	Photosensitivity, immunodeficiency, telangiectasia, developmental delay.	Skin, B-cells, T-cells, bone marrow, spleen.	[86]
Epileptic encephalopathy, early infantile, 10	613402	PNKP	SSBR	Microcephaly, seizures, developmental delay.	Brain.	[87]
Familial adenomatous polyposis-2	608456	MUTYH	BER	colorectal adenomas, adenomatous polyposis, and increased risk of colorectal cancer.	Colon, rectum, endometrium, sebaceous glands.	[88]
Immunodeficiency with hyper IgM, type 5	608106	UNG	BER	High IgM and low IgG and IgA serum concentrations.	B-cells.	[89]
Lynch Syndrome (Hereditary nonpolyposis colorectal cancer)	614385,613244 120435,609310 614337,614350	MLH1,MLH3 MSH2,MSH6 PMS1,PMS2	MMR	Increased risk of colorectal cancer, stomach, endometrial, pancreatic and urinary tract cancers.	Colon, rectum, endometrium, stomach, pancreas, urinary tract.	[90]
Spinocerebellar ataxia with axonal neuropathy	607250	TDP1	SSBR	Cerebellar ataxia, peripheral axonal motor and sensory neuropathy, gait, distal muscular atrophy.	Cerebellum, neurons, distal muscles, sural nerve.	[91]
Trichothiodystrophy	278730,601675 234050	TTDN1, TFB5 XPB, XPD	TC-NER	Brittle hair and nails, ichthyotic skin, mental retardation. Photosensitivity in half the patients.	Hair, nails, skin, neurons.	[85]
UV-sensitive syndrome	600630,614621 614640	CSB, CSB, UVSSA	TC-NER	Photosensitivity, dyspigmentation.	Skin.	[92]
Xeroderma pigmentosum	278700,610651 278720,278730 278740,278760 278780,278750	XPA, XPB, XPC XPD, XPE, XPF XPG, XPV	NER	Photosensitivity, high risk of skin carcinoma. 25% of patients have cerebral and cerebellar atrophy.	Skin, eyes, neurons, cerebrum, cerebellum.	[85]

Table 2: Mouse models for trinucleotide repeat instability discussed here.

Mouse model	Type and number* of repeats	Location of the expansion	Disease modeled	References
DM1 knock-in	84 CTGs	Knock-in of human sequences for exons 13 to 15 at the endogenous DMPK locus.	Myotonic dystrophy type 1	[35]
DM300	360 CTGs	Cosmid-based transgene containing the CTG repeat in the 3' UTR of the DMPK within a 45kb of surrounding human sequences.	Myotonic dystrophy type 1	[93]
Hdh^{Q111}	111 CAGs	Knock-in at the huntingtin locus	Huntington disease	[94]
R6/1	116 CAGs	Transgene containing the promoter and first exon of the huntingtin gene.	Huntington disease	[95]
R6/2	144 CAGs	Transgene containing the promoter and first exon of the huntingtin gene. Location of the transgene is different from the R6/1.	Huntington disease	[95]
SCA1	154 CAGs	Knock-in at the spinocerebellar ataxia type 1 locus	Spinocerebellar ataxia type 1	[96]
SCA7-CTCF-I-mut	94 CAGs	Same as RL-SCA7 92R except for a mutated CTCF binding site downstream of the repeat.	Spinocerebellar ataxia type 7	[82]
SCA7-CTCF-I-wt (also known as RL-SCA7 92R)	92 CAGs	Within a 13kb transgene of genomic DNA from the human SCA7 locus.	Spinocerebellar ataxia type 7	[97]

*: repeat number refers to the newly-generated mouse. Due to the instability, slight changes have occurred over time.

Table 3: Tissue-specificity phenotypes found in the mouse models found in table 1

Mouse model	Repeat length	BR	CB	CC	HT	KI	LI	MU	SP	ST
DM1 knock-in	84	++			–	+++	++	+		
DM300	360	++	–		–	++	+++	+		
Hdh^{Q111}	111		+	+	–	++	+++		–	+++
SCA1	154		–	++	+	++	+	+		+++
SCA7-CTCF-I-mut	94		+	++	–	+++	+++			+++
SCA7-CTCF-I-wt	92		+	++	–	+	+++			++
R6/1	116	+++	–	+	–	++	+++		–	+++
R6/2	144		–	+	–	+	+		–	+++

BR, brain; CB, cerebellum; CC, cerebral cortex; HT, heart; KI, kidney; LI, liver; MU, skeletal muscle; SC, spinal cord; SP, spleen; ST, striatum. Blank: not tested, –, no instability detected. + to +++, marginal to extreme instability. This list was updated from the one found here [98]. Only tissues with measurements in at least three different mouse models are shown.

Box 1: Quantifying TNR instability

When TNR instability was first discovered, it appeared qualitatively on Southern blots as smears that reflect the heterogeneity of the repeat length. Similarly, PCRs of an unstable TNR using large quantities of DNA produce smeared bands that are difficult to quantify. PCR-based techniques that use lower amounts of DNA can generate quantitative data on TNR instability. For instance, the Genescan approach uses a fluorescently-tagged primer to amplify about 100ng of DNA. The length of the resulting PCR fragments are then quantified using a sequencer and the Genescan software [95]. Small-pool PCR relies on amplifying about 1 to 50 molecules containing the TNR, running the products on an agarose gel before transferring the DNA onto a membrane, and then probing for the repeat locus. This approach is the gold standard and can provide precise measurements of TNR instability [99]. Other indirect methods are also widely used. They come in two flavors: plasmid-borne and chromosomal. Replication-based shuttle vectors carry an expanded repeat within a yeast reporter gene whose activity changes upon a change in TNR length [100] or the plasmid is transformed into bacteria to be amplified and the repeats are excised and run on polyacrylamide gels [101]. They have the advantage of being relatively quick to execute but are limited in the sensitivity of the assay and it is unclear how the chromatin structure on the transfected plasmids reflect the endogenous situation. The chromosomal methods rely on reporters containing CAG repeats that interfere with splicing [61, 71]. They are exquisitely sensitive, detecting contraction rates down to 10^{-5}, but they are blind to expansions.

Figure 1: Sketches of the DNA repair pathways discussed in this review. Mismatch repair (MMR) [102] is initiated when MSH2 paired with either MSH6 (to form MutSα) or MSH3 (to form MutSβ) recognizes a mismatch or a small loop, respectively. They bring in PMS2 bound to either MLH1 or MLH3, which activates the endonuclease activity of PMS2. EXO1 chews away the cut DNA containing the mismatch while RPA coats and protects the single-stranded DNA. DNA polymerase δ fills in the gap which is then ligated. TOP1-DNA adducts, base damage, and sugar damage during their processing towards repair all converge to a single-strand break (SSB) [103] and the late stages of their repair is therefore common. Sugar damage can directly generate gaps. TOP1-DNA adduct are first processed by TDP1 and PNKP to produce a SSB [103]. Other types of bulky DNA-protein crosslinks can also be processed by a combination of homologous recombination and nucleotide excision repair (NER) and can call proteases into play [104]. Base excision repair (BER) [105] is initiated by the recognition and excision of a damaged base, such as 8-oxoG, glycosylases. Monofunctional glycosylases removed the oxidized base and APEX1 creates a SSB. Bifunctional glycolylases contain a lyase activity, bypassing the need for APEX1. The resulting SSB can be filled in and ligated in either of two ways: short-patch or long-patch BER. Short patch BER involves XRCC1, LIG3, PARP1, and DNA polymerase β, which fills in one nucleotide to cover the gap. Long-patch BER involves PCNA and RFC that load one of several polymerases that synthesize at least 2 nucleotides, displacing the downstream DNA molecule. The resulting 5' flap is removed by FEN1 and ligated with LIG1. NER [75] is initiated by the recognition of bulky lesions either by the global genome repair machinery, XPC and DDB1/DDB2, or by stalling of an RNA polymerase. In the latter case, CSA, CSB, and TFIIS are involved in recruiting the downstream factors. Both of these pathways lead to the recruitment of TFIIH, which promotes melting of the DNA at the site of lesion. XPF/ERCC1 cleaves 5' of the lesion while RPA coats the template strand. PCNA and RFC can then load one of several polymerases. Strand displacement occurs and resolution and ligation are promoted by XPG, XRCC1 and LIG3, or by using FEN1 and LIG1. Colour coding: Blue, modifier of repeat instability; orange, no effect on instability; black, not tested. Other important DNA repair pathways include homologous recombination and non-homologous end-joining, which do not appear to play an important role in repeat instability in higher eukaryotes [22, 32]. Readers interested in these pathways are directed to this recent review [106].

Figure 2: Transcription and single-strand break repair in TNR instability. The expanded repeat track is indicated in blue. The molecular details of TNR instability remain unclear and thus many models have been proposed. Here are speculative models by which transcription-coupled repair (left) and single-strand break repair (right) could lead to repeat instability. Note that the RNAPII on the left would facilitate the formation of the hairpin through generation of supercoiling tension as proposed [78, 107]. In both transcription-coupled and SSBR models proposed here, the mismatch repair machinery could bind the secondary structures, stabilize them and/or affect their repair and thereby exacerbate the problem. It should be noted that these models involve intermediates in which the two DNA strands containing two different repeat lengths. These intermediates must either be repaired or replicated before leading to a symmetric duplex. For a detailed study investigating the fate of these slipped strand intermediates, see [108]. Thorough reviews of the mechanisms of TNR instability can be found here [17, 19].

References:

1 Lindahl, T. (1993) Instability and decay of the primary structure of DNA. *Nature* 362, 709-715

2 Englander, E.W. (2013) DNA damage response in peripheral nervous system: coping with cancer therapy-induced DNA lesions. *DNA Repair (Amst)* 12, 685-690

3 Iyama, T. and Wilson, D.M., 3rd (2013) DNA repair mechanisms in dividing and non-dividing cells. *DNA Repair (Amst)* 12, 620-636

4 Alexandrov, L.B.*, et al.* (2013) Signatures of mutational processes in human cancer. *Nature* 500, 415-421

5 McKinnon, P.J. (2013) Maintaining genome stability in the nervous system. *Nat Neurosci* 16, 1523-1529

6 Kohler, S.W.*, et al.* (1991) Spectra of spontaneous and mutagen-induced mutations in the lacI gene in transgenic mice. *Proc Natl Acad Sci U S A* 88, 7958-7962

7 Cervantes, R.B.*, et al.* (2002) Embryonic stem cells and somatic cells differ in mutation frequency and type. *Proc Natl Acad Sci U S A* 99, 3586-3590

8 Nouspikel, T. and Hanawalt, P.C. (2000) Terminally differentiated human neurons repair transcribed genes but display attenuated global DNA repair and modulation of repair gene expression. *Mol Cell Biol* 20, 1562-1570

9 Nouspikel, T. and Hanawalt, P.C. (2006) Impaired nucleotide excision repair upon macrophage differentiation is corrected by E1 ubiquitin-activating enzyme. *Proc Natl Acad Sci U S A* 103, 16188-16193

10 Mani, R.S. and Chinnaiyan, A.M. (2010) Triggers for genomic rearrangements: insights into genomic, cellular and environmental influences. *Nat Rev Genet* 11, 819-829

11 Karahalil, B.*, et al.* (2002) Base excision repair capacity in mitochondria and nuclei: tissue-specific variations. *FASEB J* 16, 1895-1902

12 Bell, O.*, et al.* (2011) Determinants and dynamics of genome accessibility. *Nat Rev Genet* 12, 554-564

13 Murray, J.M.*, et al.* (2012) DNA double-strand break repair within heterochromatic regions. *Biochem Soc Trans* 40, 173-178

14 Kovtun, I.V.*, et al.* (2007) OGG1 initiates age-dependent CAG trinucleotide expansion in somatic cells. *Nature* 447, 447-452

15 Nelson, D.L.*, et al.* (2013) The unstable repeats--three evolving faces of neurological disease. *Neuron* 77, 825-843

16 Gomes-Pereira, M.*, et al.* (2006) Transgenic mouse models of unstable trinucleotide repeats: toward an understanding of disease-associated repeat size mutation. In *Genetic instabilities and neurological diseases* (second edn) (Wells, R.D. and Ashizawa, T., eds), pp. 563-586, Academic Press

17 Lopez Castel, A.*, et al.* (2010) Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat Rev Mol Cell Biol* 11, 165-170

18 Mirkin, S.M. (2007) Expandable DNA repeats and human disease. *Nature* 447, 932-940

19 McMurray, C.T. (2010) Mechanisms of trinucleotide repeat instability during human development. *Nat Rev Genet* 11, 786-799

20 Cho, D.H.*, et al.* (2005) Antisense transcription and heterochromatin at the DM1 CTG repeats are constrained by CTCF. *Mol Cell* 20, 483-489

21 Kovtun, I.V.*, et al.* (2004) Somatic deletion events occur during early embryonic development and modify the extent of CAG expansion in subsequent generations. *Hum Mol Genet* 13, 3057-3068

22 Dion, V.*, et al.* (2008) Dnmt1 deficiency promotes CAG repeat expansion in the mouse germline. *Hum Mol Genet* 17, 1306-1317

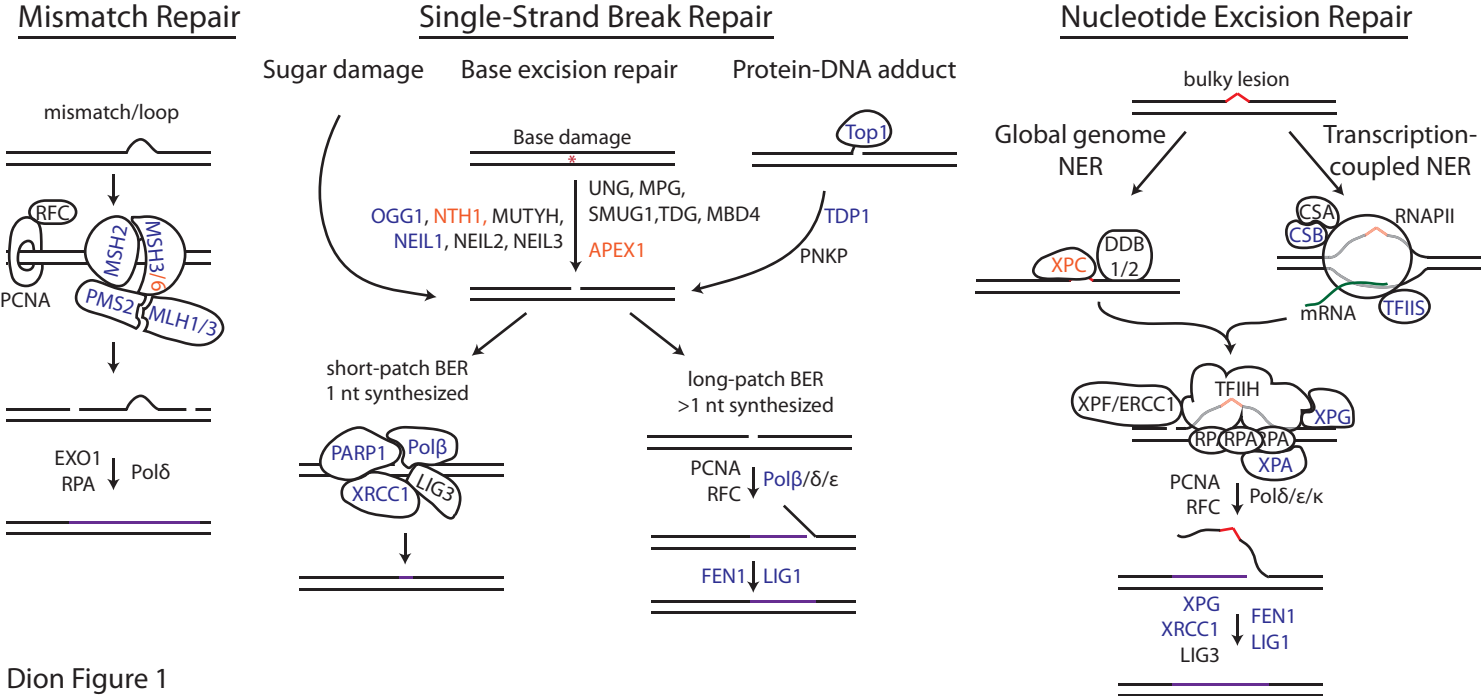
23 Cleary, J.D.*, et al.* (2010) Tissue- and age-specific DNA replication patterns at the CTG/CAG-expanded human myotonic dystrophy type 1 locus. *Nat Struct Mol Biol* 17, 1079-1087

24 Gray, S.J., *et al.* (2007) An origin of DNA replication in the promoter region of the human fragile X mental retardation (FMR1) gene. *Mol Cell Biol* 27, 426-437
25 Chastain, P.D., 2nd, *et al.* (2006) A late origin of DNA replication in the trinucleotide repeat region of the human FMR2 gene. *Cell Cycle* 5, 869-872
26 Gerhardt, J., *et al.* (2014) The DNA Replication Program Is Altered at the FMR1 Locus in Fragile X Embryonic Stem Cells. *Molecular cell* 53, 19-31
27 Mollersen, L., *et al.* (2012) Neil1 is a genetic modifier of somatic and germline CAG trinucleotide repeat instability in R6/1 mice. *Hum Mol Genet* 21, 4939-4947
28 Gomes-Pereira, M., *et al.* (2004) Pms2 is a genetic enhancer of trinucleotide CAG.CTG repeat somatic mosaicism: implications for the mechanism of triplet repeat expansion. *Hum Mol Genet* 13, 1815-1825
29 Kovalenko, M., *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
30 Kovtun, I.V. and McMurray, C.T. (2001) Trinucleotide expansion in haploid germ cells by gap repair. *Nat Genet* 27, 407-411
31 Manley, K., *et al.* (1999) Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nat Genet* 23, 471-473
32 Savouret, C., *et al.* (2003) CTG repeat instability and size variation timing in DNA repair-deficient mice. *EMBO J* 22, 2264-2273
33 Savouret, C., *et al.* (2004) MSH2-dependent germinal CTG repeat expansions are produced continuously in spermatogonia from DM1 transgenic mice. *Mol Cell Biol* 24, 629-637
34 Tome, S., *et al.* (2009) MSH2 ATPase domain mutation affects CTG*CAG repeat instability in transgenic mice. *PLoS Genet* 5, e1000482
35 van den Broek, W.J., *et al.* (2002) Somatic expansion behaviour of the (CTG)n repeat in myotonic dystrophy knock-in mice is differentially affected by Msh3 and Msh6 mismatch-repair proteins. *Hum Mol Genet* 11, 191-198
36 Wheeler, V.C., *et al.* (2003) Mismatch repair gene Msh2 modifies the timing of early disease in Hdh(Q111) striatum. *Hum Mol Genet* 12, 273-281
37 Tome, S., *et al.* (2013) MSH3 polymorphisms and protein levels affect CAG repeat instability in Huntington's disease mice. *PLoS Genet* 9, e1003280
38 Foiry, L., *et al.* (2006) Msh3 is a limiting factor in the formation of intergenerational CTG expansions in DM1 transgenic mice. *Hum Genet* 119, 520-526
39 Pinto, R.M., *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
40 Tome, S., *et al.* (2011) Maternal germline-specific effect of DNA ligase I on CTG/CAG instability. *Hum Mol Genet* 20, 2131-2143
41 Spiro, C. and McMurray, C.T. (2003) Nuclease-deficient FEN-1 blocks Rad51/BRCA1-mediated repair and causes trinucleotide repeat instability. *Mol Cell Biol* 23, 6063-6074
42 Kovtun, I.V., *et al.* (2011) Cockayne syndrome B protein antagonizes OGG1 in modulating CAG repeat length in vivo. *Aging (Albany NY)* 3, 509-514
43 Lee, J.M., *et al.* (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
44 Hubert, L., Jr., *et al.* (2011) Xpa deficiency reduces CAG trinucleotide repeat instability in neuronal tissues in a mouse model of SCA1. *Hum Mol Genet* 20, 4822-4830
45 Goula, A.V., *et al.* (2013) Tissue-dependent regulation of RNAP II dynamics: The missing link between transcription and trinucleotide repeat instability in diseases? *Transcription* 4
46 Jarem, D.A., *et al.* (2009) Structure-dependent DNA damage and repair in a trinucleotide repeat sequence. *Biochemistry* 48, 6655-6663

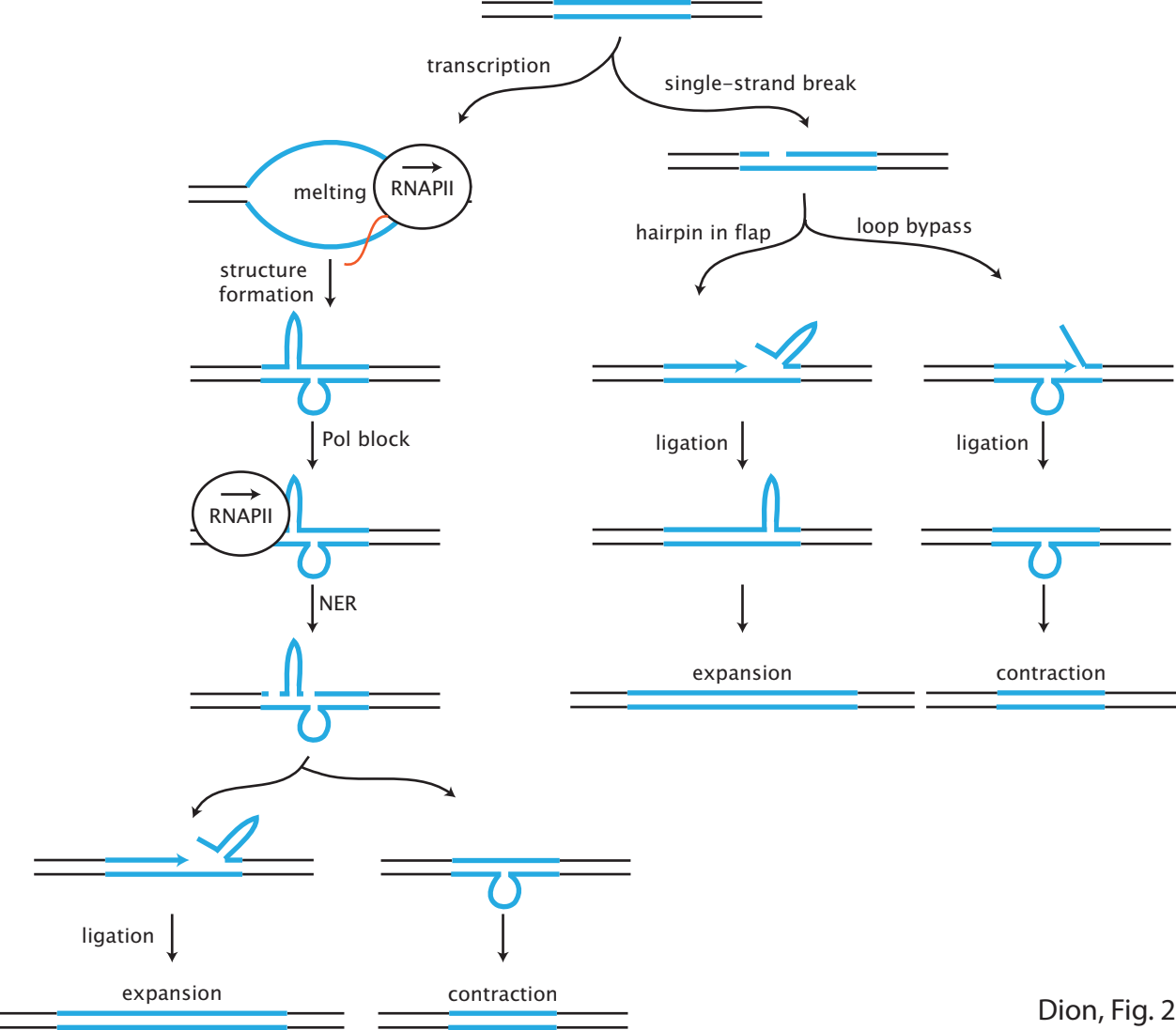
47 Goula, A.V.*, et al.* (2009) Stoichiometry of base excision repair proteins correlates with increased somatic CAG instability in striatum over cerebellum in Huntington's disease transgenic mice. *PLoS Genet* 5, e1000749
48 Imam, S.Z.*, et al.* (2006) Mitochondrial and nuclear DNA-repair capacity of various brain regions in mouse is altered in an age-dependent manner. *Neurobiol Aging* 27, 1129-1136
49 Axford, M.M.*, et al.* (2013) Detection of Slipped-DNAs at the Trinucleotide Repeats of the Myotonic Dystrophy Type I Disease Locus in Patient Tissues. *PLoS Genet* 9, e1003866
50 Lokanga, R.A.*, et al.* (2013) The Mismatch Repair Protein MSH2 is Rate Limiting for Repeat Expansion in a Fragile X Premutation Mouse Model. *Hum Mutat*
51 Bourn, R.L.*, et al.* (2012) Pms2 suppresses large expansions of the (GAA.TTC)n sequence in neuronal tissues. *PLoS One* 7, e47085
52 Ezzatizadeh, V.*, et al.* (2012) The mismatch repair system protects against intergenerational GAA repeat instability in a Friedreich ataxia mouse model. *Neurobiol Dis* 46, 165-171
53 Dragileva, E.*, et al.* (2009) Intergenerational and striatal CAG repeat instability in Huntington's disease knock-in mice involve different DNA repair genes. *Neurobiol Dis* 33, 37-47
54 Tome, S.*, et al.* (2013) Tissue-specific mismatch repair protein expression: MSH3 is higher than MSH6 in multiple mouse tissues. *DNA Repair (Amst)* 12, 46-52
55 Mason, A.G.*, et al.* (2013) Expression levels of DNA replication and repair genes predict regional somatic repeat instability in the brain but are not altered by polyglutamine disease protein expression or age. *Hum Mol Genet* doi: 10.1093/hmg/ddt551
56 Robert, T.*, et al.* (2011) HDACs link the DNA damage response, processing of double-strand breaks and autophagy. *Nature* 471, 74-79
57 Gannon, A.M.*, et al.* (2012) MutSbeta and histone deacetylase complexes promote expansions of trinucleotide repeats in human cells. *Nucleic Acids Res* 40, 10324-10333
58 Goula, A.V.*, et al.* (2012) The nucleotide sequence, DNA damage location, and protein stoichiometry influence the base excision repair outcome at CAG/CTG repeats. *Biochemistry* 51, 3919-3932
59 Oussatcheva, E.A.*, et al.* (2001) Involvement of the nucleotide excision repair protein UvrA in instability of CAG*CTG repeat sequences in Escherichia coli. *J Biol Chem* 276, 30878-30884
60 Parniewski, P.*, et al.* (1999) Nucleotide excision repair affects the stability of long transcribed (CTG*CAG) tracts in an orientation-dependent manner in Escherichia coli. *Nucleic Acids Res* 27, 616-623
61 Lin, Y.*, et al.* (2006) Transcription promotes contraction of CAG repeat tracts in human cells. *Nat Struct Mol Biol* 13, 179-180
62 Lin, Y. and Wilson, J.H. (2007) Transcription-induced CAG repeat contraction in human cells is mediated in part by transcription-coupled nucleotide excision repair. *Mol Cell Biol* 27, 6209-6217
63 Nakamori, M.*, et al.* (2011) Bidirectional transcription stimulates expansion and contraction of expanded (CTG)*(CAG) repeats. *Hum Mol Genet* 20, 580-588
64 Jung, J. and Bonini, N. (2007) CREB-binding protein modulates repeat instability in a Drosophila model for polyQ disease. *Science* 315, 1857-1859
65 Zhao, X.N. and Usdin, K. (2013) Gender and Cell-Type Specific Effects of the Transcription Coupled Repair Protein, ERCC6/CSB, on Repeat Expansion in a Mouse Model of the Fragile X-Related Disorders. *Hum Mutat*
66 Lee, J.M.*, et al.* (2010) A novel approach to investigate tissue-specific trinucleotide repeat instability. *BMC Syst Biol* 4, 29
67 Maier, T.*, et al.* (2009) Correlation of mRNA and protein in complex biological samples. *FEBS Lett* 583, 3966-3973
68 Lenzmeier, B.A. and Freudenreich, C.H. (2003) Trinucleotide repeat instability: a hairpin curve at the crossroads of replication, recombination, and repair. *Cytogenet Genome Res* 100, 7-24

69 Yang, Z*., et al.* (2003) Replication inhibitors modulate instability of an expanded trinucleotide repeat at the myotonic dystrophy type 1 disease locus in human cells. *Am J Hum Genet* 73, 1092-1105
70 Meservy, J.L*., et al.* (2003) Long CTG tracts from the myotonic dystrophy gene induce deletions and rearrangements during recombination at the APRT locus in CHO cells. *Mol Cell Biol* 23, 3152-3162
71 Gorbunova, V*., et al.* (2003) Selectable system for monitoring the instability of CTG/CAG triplet repeats in mammalian cells. *Mol Cell Biol* 23, 4485-4493
72 Lia, A.S*., et al.* (1998) Somatic instability of the CTG repeat in mice transgenic for the myotonic dystrophy region is age dependent but not correlated to the relative intertissue transcription levels and proliferative capacities. *Hum Mol Genet* 7, 1285-1291
73 Dion, V. and Wilson, J.H. (2009) Instability and chromatin structure of expanded trinucleotide repeats. *Trends Genet* 25, 288-297
74 Aguilera, A. (2002) The connection between transcription and genomic instability. *EMBO J* 21, 195-201
75 Hanawalt, P.C. and Spivak, G. (2008) Transcription-coupled DNA repair: two decades of progress and surprises. *Nat Rev Mol Cell Biol* 9, 958-970
76 Dure, L.S.t*., et al.* (1994) IT15 gene expression in fetal human brain. *Brain Res* 659, 33-41
77 Bhide, P.G*., et al.* (1996) Expression of normal and mutant huntingtin in the developing brain. *J Neurosci* 16, 5523-5535
78 Salinas-Rios, V*., et al.* (2011) DNA slip-outs cause RNA polymerase II arrest in vitro: potential implications for genetic instability. *Nucleic Acids Res* 39, 7444-7454
79 Goula, A.V*., et al.* (2012) Transcription elongation and tissue-specific somatic CAG instability. *PLoS Genet* 8, e1003051
80 Wagner, E.J. and Carpenter, P.B. (2012) Understanding the language of Lys36 methylation at histone H3. *Nat Rev Mol Cell Biol* 13, 115-126
81 Debacker, K*., et al.* (2012) Histone deacetylase complexes promote trinucleotide repeat expansions. *PLoS Biol* 10, e1001257
82 Libby, R.T*., et al.* (2008) CTCF cis-regulates trinucleotide repeat instability in an epigenetic manner: a novel basis for mutational hot spot determination. *PLoS Genet* 4, e1000257
83 Jeppesen, D.K*., et al.* (2011) DNA repair deficiency in neurodegeneration. *Prog Neurobiol* 94, 166-200
84 Date, H*., et al.* (2001) Early-onset ataxia with ocular motor apraxia and hypoalbuminemia is caused by mutations in a new HIT superfamily gene. *Nat Genet* 29, 184-188
85 DiGiovanna, J.J. and Kraemer, K.H. (2012) Shining a light on xeroderma pigmentosum. *J Invest Dermatol* 132, 785-796
86 Webster, A.D*., et al.* (1992) Growth retardation and immunodeficiency in a patient with mutations in the DNA ligase I gene. *Lancet* 339, 1508-1509
87 Shen, J*., et al.* (2010) Mutations in PNKP cause microcephaly, seizures and defects in DNA repair. *Nat Genet* 42, 245-249
88 Al-Tassan, N*., et al.* (2002) Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal tumors. *Nat Genet* 30, 227-232
89 Imai, K*., et al.* (2003) Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination. *Nat Immunol* 4, 1023-1028
90 Lynch, H.T. and de la Chapelle, A. (1999) Genetic susceptibility to non-polyposis colorectal cancer. *J Med Genet* 36, 801-818
91 Takashima, H*., et al.* (2002) Mutation of TDP1, encoding a topoisomerase I-dependent DNA damage repair enzyme, in spinocerebellar ataxia with axonal neuropathy. *Nat Genet* 32, 267-272
92 Nakazawa, Y*., et al.* (2012) Mutations in UVSSA cause UV-sensitive syndrome and impair RNA polymerase IIo processing in transcription-coupled nucleotide-excision repair. *Nat Genet* 44, 586-592

93 Seznec, H.*, et al.* (2000) Transgenic mice carrying large human genomic sequences with expanded CTG repeat mimic closely the DM CTG repeat intergenerational and somatic instability. *Hum Mol Genet* 9, 1185-1194
94 Wheeler, V.C.*, et al.* (1999) Length-dependent gametic CAG repeat instability in the Huntington's disease knock-in mouse. *Hum Mol Genet* 8, 115-122
95 Mangiarini, L.*, et al.* (1996) Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell* 87, 493-506
96 Watase, K.*, et al.* (2002) A long CAG repeat in the mouse Sca1 locus replicates SCA1 features and reveals the impact of protein solubility on selective neurodegeneration. *Neuron* 34, 905-919
97 Libby, R.T.*, 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
98 Lin, Y.*, et al.* (2006) Transcription and triplet repeat instability. In *Genetic Instabilities and Neurological Diseases* (Ashizawa, W.a., ed), pp. 691-704, Academic Press
99 Monckton, D.G.*, et al.* (1995) Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: small pool PCR analyses. *Hum Mol Genet* 4, 1-8
100 Farrell, B.T. and Lahue, R.S. (2006) CAG*CTG repeat instability in cultured human astrocytes. *Nucleic Acids Res* 34, 4495-4505
101 Cleary, J.D.*, et al.* (2002) Evidence of cis-acting factors in replication-mediated trinucleotide repeat instability in primate cells. *Nat Genet* 31, 37-46
102 Jiricny, J. (2013) Postreplicative mismatch repair. *Cold Spring Harb Perspect Biol* 5, a012633
103 Caldecott, K.W. (2008) Single-strand break repair and genetic disease. *Nat Rev Genet* 9, 619-631
104 Ide, H.*, et al.* (2011) Repair and biochemical effects of DNA-protein crosslinks. *Mutat Res* 711, 113-122
105 Krokan, H.E. and Bjoras, M. (2013) Base excision repair. *Cold Spring Harb Perspect Biol* 5, a012583
106 Kakarougkas, A. and Jeggo, P. (2013) DNA DSB repair pathway choice: an orchestrated handover mechanism. *Br J Radiol*
107 Hubert, L., Jr.*, et al.* (2011) Topoisomerase 1 and single-strand break repair modulate transcription-induced CAG repeat contraction in human cells. *Mol Cell Biol* 31, 3105-3112
108 Panigrahi, G.B.*, et al.* (2005) Slipped (CTG)*(CAG) repeats can be correctly repaired, escape repair or undergo error-prone repair. *Nat Struct Mol Biol* 12, 654-662



Dion Figure 1


transcription
single-strand break
melting
RNAPII
structure formation
hairpin in flap
loop bypass
Pol block
RNAPII
NER
ligation
ligation
expansion
contraction
ligation
expansion
contraction

Dion, Fig. 2