Stool is a sensitive and noninvasive source of DNA for monitoring expansion in repeat expansion disease mouse models

Xiaonan Zhao¹,*, Cassandra McHugh², Sydney R. Coffey², Diego Antonio Jimenez¹, Elizabeth Adams², Jeffrey B. Carroll²,*, and Karen Usdin¹,*

¹ Section on Gene Structure and Disease, Laboratory of Cell and Molecular Biology, National Institute of Diabetes, Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892
² Behavioral Neuroscience Program, Psychology Department, Western Washington University, Bellingham, WA 98225

*Corresponding authors:

Xiaonan Zhao
Gene Structure and Disease Section, Laboratory of Cell and Molecular Biology National Institute of Diabetes, Digestive and Kidney Diseases National Institutes of Health, Bethesda, MD 20892, USA
e-mail: xiaonan.zhao@nih.gov
ORCiD: 0000-0001-7281-3914

Jeff Carroll
Department of Psychology, Western Washington University
Bellingham, WA 98225, USA
e-mail: Jeff.carroll@wwu.edu
ORCiD: 0000-0003-1711-8868

Karen Usdin
Gene Structure and Disease Section Laboratory of Cell and Molecular Biology National Institute of Diabetes, Digestive and Kidney Diseases National Institutes of Health, Bethesda, MD 20892, USA
e-mail: karenu@nih.gov
ORCiD: 0000-0003-3295-0740

Keywords: Fragile X related disorder, Huntington’s disease, Spinocerebellar ataxia type 1 and 2, somatic instability, stool, peripheral DNA

Summary Statement: Stool is a readily available, non-invasive and sensitive source of DNA for monitoring repeat expansion in mouse models of four different repeat expansion diseases.
Abstract
Repeat expansion diseases (REDs) are a large group of human genetic disorders caused by expansion of a specific short tandem repeat sequence. Expansion in somatic cells affects age at onset and disease severity in some of these disorders. However, alleles in blood, a commonly used source of DNA, usually show much less expansion than disease-relevant cells in the central nervous system (CNS) in both humans and mouse models. Here we examined the extent of expansion in different DNA sources from mouse models of the Fragile X related disorders (FXDs), Huntington’s disease (HD), Spinocerebellar ataxia type 1 (SCA1) and Spinocerebellar ataxia type 2 (SCA2). We found that stool is a much better indicator of somatic expansion than blood. Since stool is a sensitive and non-invasive source of DNA, it may be useful for studies of factors affecting expansion risk or the monitoring of treatments aimed at reducing expansion in preclinical trials since it would allow expansions to be examined longitudinally in the same animal and allow significant effects to be ascertained much earlier than is possible with other DNA sources.

Introduction
More than 40+ genetic disorders are caused by expansion of a single short tandem repeat (STR) tract. This group of disorders are known as the repeat expansion diseases (REDs), and includes the Fragile X related disorders (FXDs) caused by expanded CGG/GCC repeats, and a large number of degenerative disorders caused by expanded CAG/CTG repeats such as Huntington’s disease (HD), myotonic dystrophy type 1 (DM1) and many spinocerebellar ataxias (SCAs). The length of the repeat tract that is inherited is a major determinant of the age of onset of many REDs. However, these STRs tend to increase in repeat number both on intergenerational transmission and somatically. The extent of expansion increases with time but varies with cell type, with some cell types experiencing extremely large expansions in a relatively short period of time and others remaining relatively stable for many years. Interest in somatic instability in the CAG/CTG repeat expansion diseases has been driven in part by studies showing that the tendency to expand somatically is associated with an earlier age at disease onset and/or more severe symptoms in HD (Swami et al., 2009; GeM_HD Consortium,
In addition, genome-wide association studies (GWAS) have demonstrated that variants in mismatch repair (MMR) proteins and the structure-dependent endonuclease FAN1 are associated with pronounced variation in the age of disease onset in HD mutation carriers (GeM_HD Consortium, 2015; Hong et al., 2021). Similar results have been observed in several forms of SCA caused by CAG/CTG repeat expansions (Bettencourt et al., 2016). Similarly, loss of MSH2, an important MMR protein, has been shown to modify the timing of early disease in a mouse model of HD (Wheeler et al., 2003). Since mutations in MMR genes are associated with variations in the extent of somatic expansion both in patients (GeM_HD Consortium, 2019; Bettencourt et al., 2016; Flower et al., 2019; Kim et al., 2020) and mouse models (reviewed in Zhao et al., 2021a; Wheeler and Dion, 2021), this adds weight to the idea that somatic expansion is an important modifier of disease risk and/or severity. This has led to interest in pharmacological strategies that are focused on reducing somatic instability, a strategy that may be therapeutically useful for many of these diseases (Wiggins and Feigin, 2021).

Blood is generally used for measuring the repeat size in patients. However, blood shows less somatic instability than disease-relevant cells like striatal neurons in both patients and mouse models of different repeat expansion diseases (Telenius et al., 1994; Takano et al., 1996; Mouro Pinto et al., 2020; Zhao et al., 2019; Kacher et al., 2021). A source of peripheral DNA that shows more extensive expansions would expedite testing of genetic factors affecting expansion risk and would be useful for monitoring therapeutic efforts to reduce expansions, at least in preclinical models. Previously we showed that expansion in the small intestine is more extensive than in other tissues in a FXD mouse model (Zhao et al., 2018) and is sensitive to mutations in genetic factors like FAN1 that are important modifiers of expansion risk in patient cohorts and in mice (GeM_HD Consortium, 2015; GeM_HD Consortium, 2019; Kim et al., 2020; Bettencourt et al., 2016; Ciosi et al., 2019; Mergener et al., 2020; Zhao et al., 2018; Zhao et al., 2021b). Since the murine intestinal epithelium turns over every 3-4 days (Hughes et al., 1958; Walker and Leblond, 1958), large numbers of exfoliated cells from the intestine accumulate in stool. This makes mouse stool a useful source of
genomic DNA for genotyping (Broome et al., 1999; Kalippke et al., 2009). Most of the exfoliated cells found in stool are epithelial cells from the colon (Iyengar et al., 1991). Should the colonic epithelium be as expansion prone as the cells of the small intestine, large expansions should also be detectable in stool. Here we show that the colon does show extensive expansions in mouse models of four different repeat expansion diseases and these expansions are mirrored in the DNA isolated from stool.

Results
Expansion in colon and stool is higher than most other DNA sources in an FXD mouse model.
As with many other comparable studies, we used mice with larger repeat numbers than those present in the inherited allele in some of the repeat expansion diseases. However, the available evidence suggests that the same genetic factors affect the expansion of longer repeats and shorter ones (reviewed in (Zhao et al., 2019)) and by using animals with larger repeat numbers significant expansions can be observed within an experimentally reasonable time frame.

We collected DNA samples from FXD mice with ~163 inherited repeats at 3 months of age including colon, stool as well as urine and sperm. We then and compared the extent of repeat expansion using the expansion index (EI) metric (Miller et al., 2020). The EI of tail samples taken at 3 weeks of age was used as the baseline. As shown in Fig. 1, DNA from the central nervous system (CNS), liver and tail showed a ~2-fold increase in the EI at 3 months of age. The EI in testes and sperm was somewhat higher. The similarity in the EI in these two DNA sources is consistent with our demonstration that most expansion in the testis is confined to the gametes (Zhao and Usdin, 2018b). The small intestine, distal colon and stool samples show the highest levels of expansion, with colon and stool showing very similar expansion profiles. This similarity is consistent with the fact that epithelial cells are the most common cell in the postnatal colon and are the cells most likely to be sloughed off into the intestinal lumen and thus present in stool. In contrast, as can be seen in Fig. 1B, blood and urine samples show little or no significant change in the EI relative to tail DNA at weaning.
Most DNA isolated from urine is derived from the epithelial cells from kidney and bladder (Abedini et al., 2021). Thus, it is apparent that not all epithelial cells are equally expansion prone. Since the yield of DNA from mouse urine was generally poor, we did not test this source of DNA further.

The correlation with the EI in striatum was much stronger for stool and sperm than blood (Fig. 1C). Furthermore, as can be seen in Fig. 1D, while expansions in the striatum, stool and blood increase significantly with time, the amount of expansion in blood shows relatively little change between 3 and 12 months. Thus, sperm and stool are very sensitive indicators of expansion. However, stool has the advantage that, not only is it a slightly more sensitive indicator of expansion than sperm, but it can also be used to examine expansion in both males and females.

**Expansion detected in stool can be seen at an early age and increases with age and inherited repeat size in an FXD mouse model.**

To examine the extent of repeat expansion in mouse stool, we collected fresh stool samples from FXD mice at different ages. Expansions are evident even in one month old mice with 158+ repeats, with the repeat tract in stool already being 3-4 repeats larger than it is in tail, a DNA source that shows only modest expansions (Fig. 2A). Furthermore, even at this age a modest effect of inherited allele size on the rate of repeat addition can be seen with mice with 170 repeats showing a gain of 4 repeats compared to 3 repeats for mice with 158 and 163 repeats. As can be seen in Fig. 2B, even mice with only 146 repeats have gained 3 repeats in gDNA from stool by two months of age and alleles detected in stool continue to gain repeats over a 6-month period at a relatively consistent rate resulting in the gain of 13 repeats relative to the tail DNA taken at three weeks of age. This corresponds to an average increase of 2 repeats per month. This number increases to ~3.5 repeats per month in mice with an inherited allele of 170 (Fig. 2C).
Expansion in stool reflects the effect of FAN1 in somatic expansion in an FXD mouse model.

FAN1 has been identified as a genetic modifier of disease progression in a number of repeat expansion diseases (GeM_HD Consortium, 2015; GeM_HD Consortium, 2019; Kim et al., 2020; Bettencourt et al., 2016; Ciosi et al., 2019; Mergener et al., 2020) and we previously showed that FAN1 protects against repeat expansion in FXD mouse model (Zhao and Usdin, 2018a). Furthermore, a D963A point mutation in the nuclease domain of FAN1 nuclease results in a significant increase in expansion in striatum (Zhao et al., 2021b). Here we show that the same mutation results in a significant increase in the EI in stool even in heterozygotes (Fig. 2D). In contrast, expansion in blood and sperm are similar in both WT and Fan1 mutant mice. Thus, stool reflects the effect of an important genetic modifier of expansion risk in disease-relevant cells, while blood does not.

Expansion in stool is also a good indicator of somatic expansion in CAG/CTG repeat expansion diseases.

To test whether stool is also a good indicator of somatic expansion in other repeat expansion disease models, we measured the extent of expansion in stool in a mouse model of HD. As in the FXD mice, expansions are low in heart. However, as can be seen in Fig. 3, striatum and liver are the most expansion prone tissues in this model, followed by kidney and cortex. In contrast, small intestine and distal colon have a lower EI compared with the FXD mouse. However, the EI in stool is similar to other tissues in HD mouse and higher than either blood or sperm (Fig. 3B). Significant correlations were seen between the EI in striatum and the EI in both stool and blood (Fig. 3C). However, as can be seen in Fig. 3D, the EI in stool shows larger increase with age than that seen in blood. Thus, stool is also a better indicator of somatic expansion in HD mice than blood. On average, the expansion rate in the stool of HD mice with ~112 inherited repeats is about 0.4 repeats per month (Fig. 3E).
We also examined expansion in two other CAG/CTG repeat expansion diseases, SCA1 and SCA2. As can be seen in Fig. 4, in most DNA sources the extent of expansion in SCA1 mice was similar to that seen in HD mice, with the exception of stool where expansion is significantly higher. The EI in stool was similar to that in striatum, liver, small intestine and distal colon, and higher than in other tissues including blood and sperm. Expansion in the stool of SCA1 mice with ~173 inherited repeats increases with age with ~1.5 repeats per month. In the SCA2 mice, the overall extent of expansion is much higher than in FXD, HD and SCA1 mice. As can be seen in Fig. 5, although the EI in stool is lower than striatum and liver, it is still much higher than in blood and sperm. The expansion rate in stool of SCA2 mice with ~155 inherited repeats is ~2.5 repeats per month. The high expansion in stool in both SCA1 and SCA2 mouse models suggests that expansion in stool is also a sensitive indicator of somatic expansion in the SCA1 and SCA2 mouse models.

Discussion
We have shown that repeat expansion can be detected in the stool samples of mouse models of four different repeat expansion diseases. In the FXD (Fig. 1), SCA1 (Fig. 4), and SCA2 (Fig. 5) mouse models the extent of expansion in mouse stool samples is higher than in blood and comparable to or greater to expansion in the striatum. Although expansion in stool in the HD mouse model is lower than in striatum, it is still higher than that in blood and shows a better correlation with the extent of expansion in striatum (Fig. 3). The extent of expansion in stool is comparable to that seen in distal colon consistent with the fact that most of the host DNA isolated from stool is derived from this part of the digestive system. While colonic epithelial cells are rapidly dividing and neurons are post-mitotic, available evidence from mouse models suggests that the genetic factors involved in generating somatic expansions are similar in both dividing and non-dividing cells (reviewed in (Zhao et al., 2021a)), and thus that the expansion mechanisms may also be similar. The correlation seen between the extent of expansion in stool and that seen in striatum (Fig. 1C and 3C), suggests that the extent of expansion in stool is a good indicator of the extent of expansion in the brain. However,
since expansion is not seen in urine where most host DNA is epithelial in origin, not all epithelial sources are equally prone to expansion.

Expansion in stool is apparent at an early age and is sensitive to inherited repeat size (Fig. 2). Importantly, expansion in stool also reflects the effect of an important genetic modifier of expansion risk in the FXD mouse model (Fig. 2D). We previously showed that heterozygosity for a D963A point mutation resulted in significantly more expansions in striatum, cerebellum, and liver, but not in other tissues (Zhao et al., 2021b). As can be seen in Fig. 2D, a significant increase in the extent of expansion can also been seen in the stool of these animals.

The propensity of any given cell type to expand likely reflects, at least in part, the balance between the levels of factors that promote expansion and those that protect against it. Notably, many organs show a similar extent of expansion in different RED mouse models. For example, heart shows little or no expansion in all the models, whilst the striatum and liver always show a higher level of expansion. This similarity would be consistent with the idea that the same genetic factors affect expansion in different mouse models. However, there is some discordance between the extent of expansion in some organs in different models. For example, the small intestine and distal colon show high levels of expansion in FXD, SCA1 and SCA2 mouse models, but not in the HD mouse model. Kidney shows more expansion in HD and SCA2 mouse models than either SCA1 or FXD mouse models and the extent of expansion in sperm and testes is high in the FXD model but not the others. These differences may reflect differences in the level of transcription of the affected gene or some other effect of sequence context.

Notably in all four mouse models, expansion in the CNS is highest in the striatum, even though pathology in the SCAs primarily involves the cerebellum. However, the relatively low level of expansion in cerebellum mirrors what is seen in SCA1 and SCA2 patients (Mouro Pinto et al., 2020; Chong et al., 1995; Lopes-Cendes et al., 1996; Zuhlke et al., 1997; Hashida et al., 1997; Matsuura et al., 1999). Thus, the cells that accumulate the largest expansions are not necessarily those that are most vulnerable to the
downstream consequences of expansion. This parallels the observation that tissues expressing the highest levels of the pathogenic protein are not always the sites of greatest pathology either (Sharp et al., 1995). It may be that some cells are particularly sensitive to the toxic effects of the mutant protein. In those cells, the addition of a small number of repeats could have a significant effect.

HD is the only disease in the group in which gastrointestinal (GI) dysfunction is a major symptom. Signs of this dysfunction can appear early, before evidence of CNS neurodegeneration is apparent (Kobal et al., 2018; Wood et al., 2008; Andrich et al., 2009). Similar symptoms are seen in mouse models of HD where they are associated with a decrease in the length of the colon (Stan et al., 2020), the mucosal thickness and villus length (van der Burg et al., 2011). While the mutant protein that is responsible for HD pathology is widely expressed in cells of the GI, whether dysfunction is due to GI-cell autonomous effects is unknown. Regardless, there are many other reasons why stool may be a useful source of DNA. First, stool collection is non-invasive and quick, causing minimum stress to the animals and each stool pellet provides enough mouse DNA for 10 or more PCR assays. Second, expansions in mouse stool are more extensive than in blood and are more sensitive to age and repeat size. This allows expansions to be more rapidly detected in stool from younger animals even those with smaller repeat numbers. It also reduces the time needed to see meaningful differences in the extent of expansion in mice with different genotypes or who receive different potential expansion-modifying treatments. Third, at least in the case of one important known genetic modifier of expansion risk, FAN1, expansion in stool mirrors what is seen in the brain. Finally, the simple, rapid, and non-invasive collection of stool samples allows repeat length changes to be easily and frequently monitored in the same animal over time. Thus, the use of stool DNA should expedite studies on the expansion mechanism and experimental approaches to limit these expansions. Other readily accessible sources of DNA like hair follicles may also be worth testing. However, because of low yields (Picazo and García-Olmo, 2015) and contamination risk (Cinelli et al., 2007) these may be less than ideal. Furthermore, since hair follicles from different parts of the body are comprised of cells with different embryonic origins that have
different gene expression profiles (reviewed in (Driskell et al., 2011)), they may also show differences in the extent of expansion.

There is some evidence to suggest that the propensity of different human tissues to expand is similar to that seen in these mouse models (Mouro Pinto et al., 2020; Chong et al., 1995; Lopes-Cendes et al., 1996; Zuhlke et al., 1997; Hashida et al., 1997; Matsuura et al., 1999). In addition to the higher levels of expansion in the striatum and lower levels in cerebellum of SCA1 and SCA2 patients mentioned above, HD patients also show a similar elevated level of expansion in the striatum as well as tissues like liver that also show high levels of expansion in mouse models (Mouro Pinto et al., 2020; Kennedy et al., 2003; Swami et al., 2009; Telenius et al., 1994; De Rooij et al., 1995). This raises the possibility that human stool could also show more extensive expansions than blood. While more work is required to properly evaluate the clinical use of somatic expansion as a measure of disease progression/onset risk, our results might have important clinical implications since stool may be more useful than blood for assessment of somatic expansion risk. Stool may also be useful for monitoring the efficacy of clinical trials of therapies aimed at reducing somatic expansion. Furthermore, the similarity we have seen in the cell type specificity of expansion in the four disease models suggests that stool may be a useful peripheral source of DNA for monitoring expansions in patients with other repeat expansion diseases and/or their mouse models.

**Materials and Methods**

**Reagents**

All reagents were from Sigma-Aldrich (St. Louis, MO) unless otherwise specified. Primers were from Life Technologies (Grand Island, NY). Capillary electrophoresis of fluorescently labeled PCR products was carried out by the Roy J Carver Biotechnology Center, University of Illinois (Urbana, IL).
Mouse generation, breeding and maintenance

The generation of FXD and Fan1 D963A mice was described previously (Entezam et al., 2007; Zhao et al., 2021b). These mice were maintained at NIH in a manner consistent with the Guide for the Care and Use of Laboratory Animals (NIH publication no. 85-23, revised 1996) and in accordance with the guidelines of the NIDDK Animal Care and Use Committee, who approved this research (ASP-K021-LMCB). The generation of the HD (Wheeler et al., 1999), SCA1 (Watase et al., 2002), and SCA2 (Sen et al., 2019) mouse models were described previously. HD mice, SCA1 mice (provided by Huda Zoghbi, Howard Hughes Medical Institute, Baylor College of Medicine) and SCA2 mice (provided by Georg Auburger, Goethe University Medical School, Frankfurt), were maintained at Western Washington University (WWU) in a manner consistent with protocols approved by the WWU institutional animal care and use committee. All mice are on a C57BL/6J background.

Mouse stool and urine sample collection

For stool collection mice were moved to a clean cage with a mat. Three to five pieces of fresh stool were collected in a 1.5 mL tube and transferred to dry ice. Stool samples were kept at -80˚C until further processing. Mouse urine was collected in one of two ways. The first involved holding the animal over a 1.5 mL collection tube while lightly stroking its belly or by placing the mouse in a clean, dry, empty cage covered with a plastic wrap until it urinates. The urine was then aspirated with a pipette and transfer to the collection tube. These steps were repeated until at least 300 μL urine sample was collected. Urine samples were kept at 4˚C for up to 24 hours before processing, or at -80˚C until further processing.

DNA isolation

DNA from mouse tails at 3-weeks-old was extracted for genotyping using the KAPA Mouse Genotyping Kit (KAPA Biosystems, Wilmington, MA). A 5 cm region of the jejunum starting 10 cm downstream of stomach was used as the small intestine sample. A 5 cm region of the colon upstream of anus was used as the distal colon sample. DNA from tissue samples was isolated using a Maxwell®16 Mouse Tail DNA Purification Kit.
(Promega, Madison, WI) according to the manufacturer’s instructions. DNA from blood was isolated using a Maxwell® 16 Blood DNA Purification Kit (Promega, Madison, WI) according to the manufacturer’s instructions. Sperm collection and gDNA preparation as previously described (Zhao and Usdin, 2018b). DNA from stool was isolated using Norgen Stool DNA Isolation Kit (Norgen Biotek, Thorold, Ontario, Canada) according to the manufacturer’s instructions. A single mouse stool pellet weighs an average of 30.8 mg and yields ~5.5 µg of DNA. This is sufficient for >10 repeat PCR assays (see below). DNA from urine was isolated using Norgen Urine DNA Isolation Micro Kit (Norgen Biotek, Ontario, Canada) according to the manufacturer’s instructions.

**Genotyping and analysis of repeat number**

Repeat size analysis of the Fmr1, Atxn1, Atxn2, and Htt alleles in the FXD, SCA1, SCA2, and HD mice respectively was carried out using a fluorescent PCR assay with FAM-labeled primer pairs. The primers FAM-labelled FraxM4 (FAM-5’-CTTGAGGCCCAGCCGCCGTCGGGCC-3’) and FraxM5 (5’-CGGGGGGGCTGCGGTAACGGCCCAA-3’) were used for the Fmr1 allele (Entezam et al., 2007), the primers FAM-labelled 8930 (FAM-5’-CAGACGCGGACACAAG-3’) and 8931 (5’-ATCATCGTCTGATGGGGATG-3’) were used for the Atxn1 allele (Watase et al., 2002), the primers FAM-labelled SCA2Ex1-Fwd5 (FAM-5’-CCCGGCGGCGTGAGCGCCGTTGAT-3’) and SCA2Ex1-Rev2 (5’-CGGGCTTGCGCCAGTGG-3’) were used for the Atxn2 allele (Sen et al., 2019), and the primers FAM-labelled CAG1 (FAM-5’-ATGAAGGCCTTCGAGTCCCTCAAGTCCTTC-3’) and HU3 (5’-GGCGGCTGAGGAGCTGAGGA-3’) were using for the Htt allele (Lee et al., 2010).

The amount of gDNA template in the PCR mix varied for different samples. For blood, sperm, and tissue samples, 100 ng gDNA was used as template. For stool sample, 200-400 ng gDNA was used as template since gDNA isolated from stool contains a large amount of microbial DNA. For urine samples, 10-50 ng gDNA was used as template since the gDNA yield varied between urine samples from different animals. The PCR mix for Fmr1, Atxn1, and Atxn2 allele contained 2 µL DNA template, 1x KAPA2G Fast HotStart Genotyping Mix (KAPA Biosystems, Wilmington, MA), 2.4 M Betaine, 2%
DMSO, and 0.5 μM each of the primers. PCR parameters for the Fmr1 allele were 95°C for 10 min, 35 cycles of (95°C for 30 s, 65°C for 30 s, and 72°C for 90 s), followed by incubation 72°C for 10 min. PCR parameters for the Atxn1 allele were 95°C for 10 min, 40 cycles of (95°C for 30 s, 58°C for 30 s, and 72°C for 90 s), followed by incubation 72°C for 10 min. PCR parameters for the Atxn2 allele were 95 °C for 10 min, 40 cycles of (95°C for 40 s, 60°C for 40 s, and 72°C for 90 s), followed by incubation 72°C for 10 min. The PCR mix for Htt allele contained 2 μL DNA template, 1× KAPA2G Fast HotStart Genotyping Mix (KAPA Biosystems, Wilmington, MA), 1.2 M Betaine, 1% DMSO, and 0.5 μM each of the primers. Touchdown PCR was used to amplify the Htt allele with the following parameters: 95 °C for 10 min, 10 cycles of (95°C for 30 s, 72°C with -1°C/cycle for 30 s, and 72°C for 90 s), 28 cycles of (95°C for 30 s, 63°C for 30 s, and 72°C for 90 s), followed by incubation 72°C for 10 min. The PCR products were resolved by capillary electrophoresis on an ABI Genetic Analyzer. The resultant fsa file was then displayed using a previously described custom R script (Hayward et al., 2016) that is available on request.

The number of repeats in the modal allele found in tail samples collected at 3 weeks old was used as an indicator of the number of original inherited repeats. For tissues showing a unimodal distribution of allele sizes, the difference between the repeat number present in the modal allele and the repeat number of the modal allele in tail at 3 weeks old was used as a measure of the extent of expansion. For tissues with a bimodal distribution of alleles, the extent of expansion was calculated by subtracting the repeat number in tail at 3 weeks old from the number of repeats of the modal allele in the larger of the two allele populations. The expansion rate was calculated based on data from an average of at least two different time points between 2- to 12-months from the same animal. We also quantified somatic expansions using the expansion index (EI) as a metric (Miller et al., 2020). The EI of tail sample taken at 3 weeks old was used as the baseline.
Statistical analyses

Statistical analyses were performed using GraphPad Prism 9.3. For comparisons of EI in different samples to EI in stool or tail DNA taken at weaning, statistical significance was assessed using either a mixed-effects model when not all organs were available for all animals, or a repeated measures (RM) one-way ANOVA, both with Geisser-Greenhouse correction and Dunnett’s correction for multiple comparisons. For comparisons of EI in samples with different age or different genotype, statistical significance was assessed using the RM two-way ANOVA with Geisser-Greenhouse correction and Tukey’s correction for multiple comparisons. For comparisons of EI in stool and blood at different age, statistical significance was assessed using paired (stool vs blood) or unpaired (6-month vs 14-month) two-tailed t test with Holm-Sidak’s correction for multiple comparisons. Correlation between EI of different tissues was assessed using liner regression.

Acknowledgements

The authors would like to thank Professors Georg Auburger and Huda Zoghbi for kindly providing us with the SCA2 and SCA1 mice, respectively. We also would like to acknowledge all the hard work done by the staff who take care of our mice and without whom this work would not have been possible.

Competing interests

Jeffrey B. Carroll is on the scientific advisory board, and have shares, in Triplet Therapeutics, which is a company focused on modulating somatic instability in HD and other diseases. Other authors declare no competing or financial interests.

Author contributions

Funding statement

This work was made possible by funding from the Intramural Program of the National Institute of Diabetes, Kidney and Digestive Diseases to KU (DK057808). Work in the Carroll lab was supported by funds provided by the Behavioral Neuroscience Program at Western Washington University.

References


Figures

A  FXD mouse model

B  CNS  peripheral tissues  other DNA sources

C  Striatum expansion index

D  Expansion index
Fig. 1. Quantitative analyses of CGG repeat expansions in FXD mice. A) Typical repeat PCR profiles from tail taken at 3 weeks old (3 wk) and different sources of DNA from a 3-month-old FXD mouse with 163 repeats. The dotted line represents the size of the original inherited allele as ascertained from the tail DNA taken at 3 weeks old. B) Comparison of the expansion index in different organs and DNA sources of 3-month-old FXD mice with an average of 163 repeats in the original allele. The lower dotted line represents the basal expansion level as ascertained from the tail DNA taken at 3 weeks old. The upper dotted line represents the expansion level in stool. The data represent the average of four male mice with 160-164 repeats. The error bars indicate the standard deviations of the mean. Each dot represents one animal. The EI in different DNA sources were compared to the EI in stool and tail DNA taken at weaning using a repeat measures (RM) one-way ANOVA with correction for multiple testing as described in the Materials and Methods. The adjusted $P$ values levels are listed in the table below. *: $P<0.05$; **: $P<0.01$; ***: $P<0.001$; ns: not significant. C) Correlation between EI in striatum and EI in stool, sperm, and blood of 12 male FXD mice with 158-164 repeats at different ages. D) Boxplot of the EI of male FXD mice at different ages. Each age group including four animals with 158-164 repeats. Each dot represents one animal. The significance of the age effect was assessed using RM two-way ANOVA with correction for multiple testing as described in the Materials and Methods. **: $P<0.01$; ***: $P<0.001$; ****: $P<0.0001$; ns: not significant.
Fig. 2. Quantitative analyses of CGG repeat expansions in stool from FXD mice of different inherited allele sizes and genetic conditions. A) Typical repeat PCR profiles from tail taken at 3 weeks old (3 wk), and stool samples collect from FXD mice with different inherited repeat numbers at one month old (1 mo). The numbers associated with some of the profiles indicate the number of repeats added during the lifetime of the mouse. The dotted line represents the size of the original inherited allele as ascertained from the tail DNA taken at 3 weeks old. B) Typical repeat PCR profiles from tail taken at 3 weeks old, and stool samples collected from the same FXD mouse over time. The numbers associated with the profiles indicate the number of repeats...
added during the lifetime of the mouse. The dotted line represents the size of the original inherited allele as ascertained from the tail DNA taken at 3 weeks old. C) Repeats added per month in male FXD mice with different repeat numbers, including five mice with 145-146 repeats, eight mice with 153-160 repeats, nine mice with 162-166 repeats, and four mice with 167-173 repeats. Each dot represents one animal. The repeat size range and average repeat size are listed in the table below. D) Boxplot of the EI in striatum and other sources DNA from 6-month-old FAN1 WT and FAN1 D963A mutant mice with an average of 161 repeats in the original allele. The data for each genotype is based on four animals with 159-164 repeats. Each dot represents one animal. The significance of the genotype effects was assessed using RM two-way ANOVA with correction for multiple testing as described in the Materials and Methods. *: P<0.05; ***: P<0.001; ****: P<0.0001; ns: not significant.
Fig. 3. Quantitative analyses of CAG expansions in HD mice. A) Typical repeat PCR profiles from tail taken at 3 weeks old (3 wk) and samples from different sources of DNA from 6-month-old HD mice with 109 repeats. The dotted line represents the size of the...
original inherited allele as ascertained from the tail DNA taken at 3 weeks old. B) Comparison of the EI in different organs and DNA sources of 6-month-old HD mice with an average of 114 repeats in the original allele. The lower dotted line represents the basal expansion level as ascertained from the tail DNA taken at 3 weeks old. The upper dotted line represents the expansion level in stool. Testes and sperm samples represent the average of five male mice with 109-124 repeats. Other data represents the average of five male and three female mice in the same repeat range. The error bars indicate the standard deviation of the mean. Each dot represents one animal. The EI in different DNA sources were compared to the EI in stool and in tail DNA taken at weaning using a mixed-effects model with correction for multiple testing as described in the Materials and Methods. The adjusted $P$ values levels are listed in the table below. *: $P<0.05$; **: $P<0.01$; ***: $P<0.001$; ****: $P<0.0001$; ns: not significant. C) Correlation between EI in striatum and EI in stool or blood of 12 HD mice with 109-124 repeats at different ages. D) Boxplot of the EI in stool and blood sample in C), which were collected from five male and three female HD mice with 109-124 repeats at 6 months old, and four male HD mice with 111-113 repeats at 14 months old. Each dot represents one animal. The significance was assessed using paired (stool vs blood) or unpaired (6-month vs 14-month) t test with correction for multiple testing as described in the Materials and Methods. **: $P<0.01$; ***: $P<0.001$; ****: $P<0.0001$. E) Boxplot of the repeat added in stool sample collected from HD mice with an average of 112 repeats at different ages, including nine male mice at 3 months old with 110-116 repeats, five male and three female mice at 6 months old with 109-124 repeats, six male mice at 10 months old with 108-116 repeats, and four male mice at 14 months old with 111-113 repeats. Each dot represents one animal.
Fig. 4. Quantitative analyses of CAG expansions in SCA1 mice. A) Typical repeat PCR profiles from tail taken at 3 weeks old (3 wk) and samples from different sources of DNA from 4-month-old SCA1 mice with 173 repeats. The dotted line represents the size of the original inherited allele as ascertained from the tail DNA taken at 3 weeks old. B) Comparison of the expansion index in different organs and DNA sources of 4-month-old SCA1 mice with an average of 171 repeats in the original allele. The lower dotted line represents the basal expansion level as ascertained from the tail DNA taken at 3 weeks old. The upper dotted line represents the expansion level in stool. Testes and sperm samples represent the average of two male mice with 171 and 173 repeats. Other data represents the average of two male and one female mice in the same repeat range. The
error bars indicate the standard deviations of the mean. Each dot represents one animal. The EI in different DNA sources were compared to the EI in stool and tail DNA taken at weaning using a mixed-effects model with correction for multiple testing as described in the Materials and Methods. The adjusted $P$ values levels are listed in the table below. *: $P<0.05$; **: $P<0.01$; ns: not significant. C) Boxplot of the repeat added in stool sample collected at different age from the same animals, including two male and two female SCA1 mice with 170-176 repeats. Each dot represents one animal. Only three animals were available at 4 months of age.
Fig. 5. Quantitative analyses of CAG expansions in SCA2 mice. A) Typical repeat PCR profiles from tail taken at 3 weeks old (3 wk) and samples from different sources of DNA from 4-month-old SCA2 mice with 160 repeats. The dotted line represents the size of the original inherited allele as ascertained from the tail DNA taken at 3 weeks old. B) Comparison of the expansion index in different organs and DNA sources of 4-month-old SCA2 mice with an average of 158 repeats in the original allele. The lower dotted line represents the basal expansion level as ascertained from the tail DNA taken at 3 weeks old. The upper dotted line represents the expansion level in stool. Testes and sperm samples represent the average of four male mice with 154-162 repeats. Other data represents the average of four male and two female mice in the same repeat range. The
error bars indicate the standard deviations of the mean. Each dot represents one animal. The EI in different DNA sources were compared to the EI in stool and tail DNA taken at weaning using a mixed-effects model with correction for multiple testing as described in the Materials and Methods. The adjusted $P$ values levels are listed in the table below. *: $P<0.05$; **: $P<0.01$; ***: $P<0.001$; ****: $P<0.0001$; ns: not significant. C) Boxplot of the repeat added in stool sample collected at different age from the same animals including four male and two female SCA2 mice with 154-162 repeats at different age. Each dot represents one animal.