Supplementary Material

Somatic CAG Repeat Stability in a Transgenic Sheep Model of Huntington's Disease

Supplementary Methods 1: 5-year-old OVT73 Small Pool PCR

PCR primers

Sheep outer F: 5'-TAGGGCTGTCAATCATGCTG -3'

Sheep Outer R: 5'-CACGGTCTTTCTTGGTAGCTG-3'

Laragen F 6-FAM: 5' 6-FAM-CGGCCGCTCAGGTTCTGCTTTTACCTG -3'

Laragen R: 5'-AGCAGCGGCTGTGCCTGCGG-3'

Laragen F NED: 5' NED-CGGCCGCTCAGGTTCTGCTTTTACCTG -3'

Small pool PCR (SP-PCR)

Genomic DNA was extracted from ~30 mg fresh-frozen liver and striatum tissue dissected post-mortem from a cohort of six transgenic OVT73 sheep aged five years, using a standard phenol-chloroform method [1]. Tissue samples were obtained from the South Australian Research and Development Institute (SARDI) and were sampled in accordance with approval of the Department of Primary Industries and Regions (PIRSA) Animal Ethics Committee (Approval number 19/02). DNA was digested with *HindIII* restriction enzyme (Roche Diagnostics, Germany) and serially diluted in 1x Tris-EDTA buffer containing 0.1 µM of Sheep Outer F primer. DNA equivalent to a single amplifiable allele per μL was determined empirically using Poisson analysis and used for repeat length typing following nested PCR, prepared using a KAPA2G Robust PCR Kit (KAPA Biosystems, USA).

First round amplification was carried out using 1 genomic equivalence with primers Sheep Outer F and Sheep Outer R. Second round PCR was carried out using 1:100 dilution (in PCR-grade water) of the first round PCR product and amplified using primers; Laragen F 6-FAM and Laragen R. The forward primer was 6-FAM labelled to enable resolution of the PCR product and accurate quantification of CAG repeat length using an automated ABI3130XL DNA sequencer with GeneScan™ 600 LIZ® Size Standard (Applied Biosystems). Primers were specific to the transgene (human *HTT*) and did not amplify wild-type sheep *HTT*. For both PCR reactions, mixtures (10 µL) contained 0.5 µM each primer, 0.2 mM dNTP Mix, 0.5 U KAPA 2G Robust, 5%

volume DMSO, GC KAPA buffer, and PCR-grade water. Mixtures were prepared in bulk and aliquoted across 96-well PCR plates before adding 1 µL of the appropriate template. Cycling conditions were 94°C for 10 min, 30x cycles [94°C for 30 s, 58°C for 30 s, and 72°C for 60 s] and a final extension at 72°C for 7 min. All PCR set up and template preparation was performed in a hood, with segregated pipettes, plasticware and equipment. Sequencing and GeneScan analysis of SP-PCR products was performed at the University of Auckland Centre for Genomics, Proteomics and Metabolomics (University of Auckland).

Quality control samples for SP-PCR

Every 96 well plate contained 16 no-template negative controls. A positive control sample for which the CAG repeat was previously determined by bulk PCR was also loaded onto every 96-well PCR plate to facilitate accurate repeat length sizing of samples. Different approaches were used for striatum and liver as follows. For the assessment of striatum samples, a positive control template (bulk tail DNA from sheep sample HD383, previously bulk genotyped at (CAG)69(CAACAG)2(CCG)9(CCT)3) was added in triplicate to each 96-well PCR plate, with reaction mixtures prepared as for the samples. For the assessment of liver samples, a previously prepared 5'NED-labelled second round amplification product for a positive control template (single genome equivalence of tail DNA from sheep sample HD372, PCR product previously bulk genotyped with (CAG)69(CAACAG)^2(CCG)9(CCT)3) was spiked-in to every sample well on each 96-well PCR plate as the plates were prepared for Genescan analysis. The use of a 5' NED-labelled primer Laragen F NED as the second-round amplification forward primer was to enable the spiked-in control amplification product to be distinguished from sample amplification products in the same well (5' FAM labelled).

GeneScan analysis for SP-PCR

One microlitre of each second round PCR product was mixed with 0.5 µL GeneScan 600LIZ Size Standard, 9.5 µL Formamide (HiDi Formamide, Applied Biosystems, USA). For liver samples, 0.2 µL of the NED labelled spike-in positive control was also added to each well at this point. Mixtures were heat denatured for 3 min and cooled on ice. Fragment analysis was performed within two days of amplification, on an ABI3130XL DNA sequencer at the Centre for Genomics,

Proteomics and Metabolomics Centre within the School of Biological Sciences, University of Auckland.

Resulting .fsa files were analysed using the Geneious microsatellite plugin (Genieous 6.1.4, Biomatters Ltd, NZ) to enable peak sizing and calling against the LIZ600 ladder. The PCR products obtained from single molecule inputs comprise a cluster of peaks, each differing by one CAG repeat, due to PCR stutter. Allele size was assigned to be the highest peak in the cluster. Pure CAG repeat length was then estimated by comparison to the highest peak of the positive control with known transgene structure (CAG)69(CAACAG)2(CCG)9(CCT)3, according to the following equation:

$$Number\ of\ repeats\ = \left(\frac{SP - CP}{3}\right) + KL$$

where SP is Sample Peak length in base pairs (from Genescan), CP is the positive Control Peak length in base pairs (from Genescan) and KL is the Known Length of the positive control (69 pure CAG).

Data was used only from plates in which the following quality control conditions were met: 1) all negative control wells showed no peak, 2) positive control peaks were present, and 3) Poisson distribution was satisfied (at least 36% of samples returned no peak). This was done until at least 50 molecules of data were collected for each animal. The inclusion of the spike-in control in every sample well for the liver sample assessment assisted with the accuracy of bin assignments of called peaks.

Supplementary Methods 2: OVT73 founder (HD260) analysis

Bulk PCR of tail tissue derived from the OVT73 founder animal (HD260)

Tail tissue from OVT73 founder HD260 and five other founder animals (HD185, HD208, HD227, HD261, HD287) was obtained from the South Australian Research and Development Institute (SARDI) in accordance with approval of the Department of Primary Industries and Regions (PIRSA) Animal Ethics Committee (Approval number 19/02). Genomic DNA was extracted from fresh-frozen tail tissue biopsy and RNase A treated using a Qiagen DNeasy® Blood and Tissue kit and QIAGEN TissueLyser according to the manufacturer's guidelines (QIAGEN, Germary). PCR was performed using Laragen F and Laragen R primers described in Supplementary Methods 1, to confirm the presence of the OVT73 transgene. PCRs were set up

using the Expand High Fidelity PCR system (Roche). Each reaction contained 20 ng genomic DNA, 0.5 µM each primer, 10% DMSO and 0.35 U ThermaGo polymerase. Cycling conditions were 94°C for 5 min, followed by 32 x cycles [94°C for 30 s, 60°C for 30 s, and 72°C for 1 min 30 s] and a final extension at 72°C for 7 min. PCR products were resolved by gel electrophoresis through a 1% agarose gel, with size estimation by inclusion of 1 Kb Plus ladder (Invitrogen).

Generation of OVT73 founder (HD260) fibroblasts

Fibroblasts were derived from tail tissue explants of animal HD260 and five other potential founder animals. Cells were maintained at 38.5°C and 5% CO_2 in DMEM high glucose media (Invitrogen) supplemented with 10% fetal bovine serum, penicillin-streptomycin (each at 100 units per mL), gentamicin (0.005 mg/mL), and fungizone 5 (0.25 µg/mL) (all from Invitrogen).

DNA and RNA extraction from OVT73 founder (HD260) fibroblasts

Fibroblast cells were harvested (0.25% Trypsin 0.03% EDTA, followed by PBS washes and centrifugation to pellet cells) and DNA and RNA extracted from cell pellets. For DNA the pelleted cells were lysed in 500 µL Lysis Buffer (100 mM Tris pH 8.0, 5 mM EDTA, 200 mM NaCl, 0.2% SDS, 0.1 mg/mL Proteinase K) and DNA then precipitated by incubation in ice cold isopropanol for 1hr followed by centrifugation (12,000 g at 4°C for 20 min) and washing in 70% ethanol. After washing and centrifugation (12,000 g at 4°C for 20 mins) the DNA pellet was air dried and resuspended in 20 µL Tris EDTA buffer (Tris 10 mM, EDTA 1 mM, pH 8.0). DNA was RNase A treated using a Gentra® Puregene® Cell Kit according to the manufacturer's instructions (v 09/2007) (Qiagen, Germany). RNA was extracted using a TRIzol method. The cell pellet was mixed with 1 mL TRIzol reagent (Invitrogen) and incubated for 5 min at room temperature. Chloroform (0.2 mL) was then added, and the tube shaken vigorously for 15 s followed by incubation at room temperature for 2-3 min. Subsequently the mixture was centrifuged at 12,000 g for 15 min at 4°C and the upper aqueous phase containing RNA removed to a new tube. Isopropanol (0.5 mL) was added to precipitate the RNA. The tube was inverted 10 times and incubated at room temperature for 10 min, then centrifuged at 12,000 g for 10 min at 4°C. The pellet was washed in 75% ethanol followed by centrifugation at 7,500 g for 5 min at 4°C. The pellet was then air dried before resuspension of RNA in 50 µL distilled water. Quantification and integrity of DNA and RNA was assessed by Nanodrop (NanoDrop Technologies).

Bulk PCR of complementary DNA derived from OVT73 fibroblasts

PCR was performed on complementary cDNA derived from RNA to examine transcription of the transgene. Synthesis of cDNA was performed using the SuperScript® III First-Strand Synthesis System for RT-PCR (Invitrogen) according to the manufacturer's instructions, using 500 ng RNA. A reaction lacking superscript enzyme was also generated for each sample as a negative control. PCR was then performed on 1 μL cDNA using Laragen F and Laragen R primers and cycling conditions as described for the corresponding genomic DNA PCR. PCR products were resolved by gel electrophoresis through a 1% agarose gel, with size estimation by inclusion of 1kb Plus ladder (Invitrogen).

Small pool PCR analysis of OVT73 founder (HD260) fibroblasts

Small pool PCR was performed using 1 genomic equivalence using the method described for SP-PCR in Supplementary Methods 1. CAG repeat sizes were resolved using an automated ABI3730XL DNA Analyzer and GeneMapper v.3.7 software with GS 500 LIZ internal size standard (Applied Biosystems), at the MGH Mission Driven Service Core for DNA Fragment Analysis. Control samples of known *HD* CAG repeat length were included in every run. Allele size was assigned to the highest peak in the cluster.

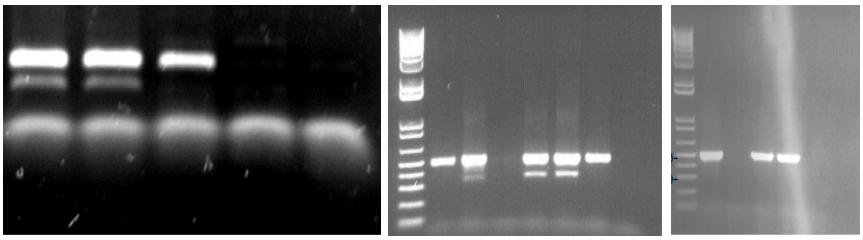
Supplementary Methods 3: 10-year-old OVT73 bulk PCR analysis

Genomic DNA was isolated using the DNeasy DNA blood and tissue kit (Qiagen). Somatic CAG instability analysis was performed using a human-specific PCR assay (Forward primer 5' ATGAAGGCCTTCGAGTCCCTCAAGTCCTTC; reverse primer 5' GGCGGCTGAGGAAGCTGAGGA) that amplifies the *HTT* CAG repeat from the OVT73 transgene [2]. The forward primer was fluorescently labelled with 6-FAM (Applied Biosystems; ABI) and products were resolved using the ABI 3730xl DNA analyzer (ABI) with GeneScan 500 LIZ as internal size standard (ABI). CAG repeat size was estimated in GeneMapper v5 against knock-in mouse model standards with the human canonical polyglutamine-polyproline repeat sequence structure (CAG)nCAACAGCCGCCA(CCG)7(CCT)2 and then adjusted based on the structure in the sheep. MiSeq-based sequencing of OVT73 transgene [4] showed the polyglutamine-polyproline repeat structure to be (CAG)nCAACAGCAACAG(CCG)9(CCT)3

(Supplementary Figure 4). As the reverse primer in the PCR sizing assay includes this entire sequence (the 3' end of the reverse primer anneals to TCCT that encompasses the ultimate CCT triplet), the OVT73 transgene amplicon will be 9 bp (3 triplets) larger than the KI mouse standard amplicon for the same number of pure CAG repeats, due to the presence of an additional CAACAG sequence and an additional CCT triplet in the sheep. To account for this, pure CAG length in the sheep was estimated to be 3 repeats less than that assumed based on the KI mouse repeat. It is worth noting that that fragment sizing assumes that single stranded fragments of the same base-pair length have the same electrophoretic mobility, which may not be the case for the sheep alleles and mouse knock-in standards that have different sequence structures and nucleotide composition. We also note that the reverse primer may mis-prime in the sheep transgene due to the additional CCT triplet.

Disclosure of uncropped /raw agarose gel images in this manuscript

The images shown below are the raw gel electrophoresis images relating to Supplementary Figure 1, panels B and C respectively. Images were cropped for the manuscript figure.



REFERENCES

[1] Sambrook J, Fritsch EF, Maniatis T. Extraction with phenol: chloroform. In: Sambrook J, Fritsch EF, Maniatis T, edits. Molecular cloning: a laboratory manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1989.

[2] Pinto RM, Dragileva E, Kirby A, Lloret A, Lopez E, St. Claire J, et al. Mismatch repair genes Mlh1 and Mlh3 modify CAG instability in Huntington's disease mice: genome-wide and candidate approaches. PLoS Genet. 2013;9(10):e100393.

[3] Lee JM, Zhang J, Su AI, Walker JR, Wiltshire T, Kang K, et al. A novel approach to investigate tissue-specific trinucleotide repeat instability. BMC Syst Biol. 2010 Mar 19;4(1):29.

[4] Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium. CAG repeat not polyglutamine length determines timing of Huntington's disease onset. Cell. 2019;178(4): 887-900.e14.

Supplementary Table 1. Description of OVT73 sheep samples assessed for CAG repeat instability.

Sample	Transgenic status	Sex	Age (y,m)	Generation	Analysis Method
HD372	Transgenic	Ewe	5y, 4m	G2	SP-PCR
HD377	Transgenic	Ewe	5y, 4m	G2	SP-PCR
HD376	Transgenic	Ewe	5y, 4m	G2	SP-PCR
HD317	Transgenic	Ram	5y, 10m	G1	SP-PCR
HD339	Transgenic	Ram	5y, 10m	G1	SP-PCR
HD383	Transgenic	Ram	5y, 4m	G2	SP-PCR
HD909	Transgenic	Ewe	10y	G3	Bulk PCR
HD912	Transgenic	Ewe	10y	G3	Bulk PCR
HD913	Transgenic	Ewe	10y	G3	Bulk PCR

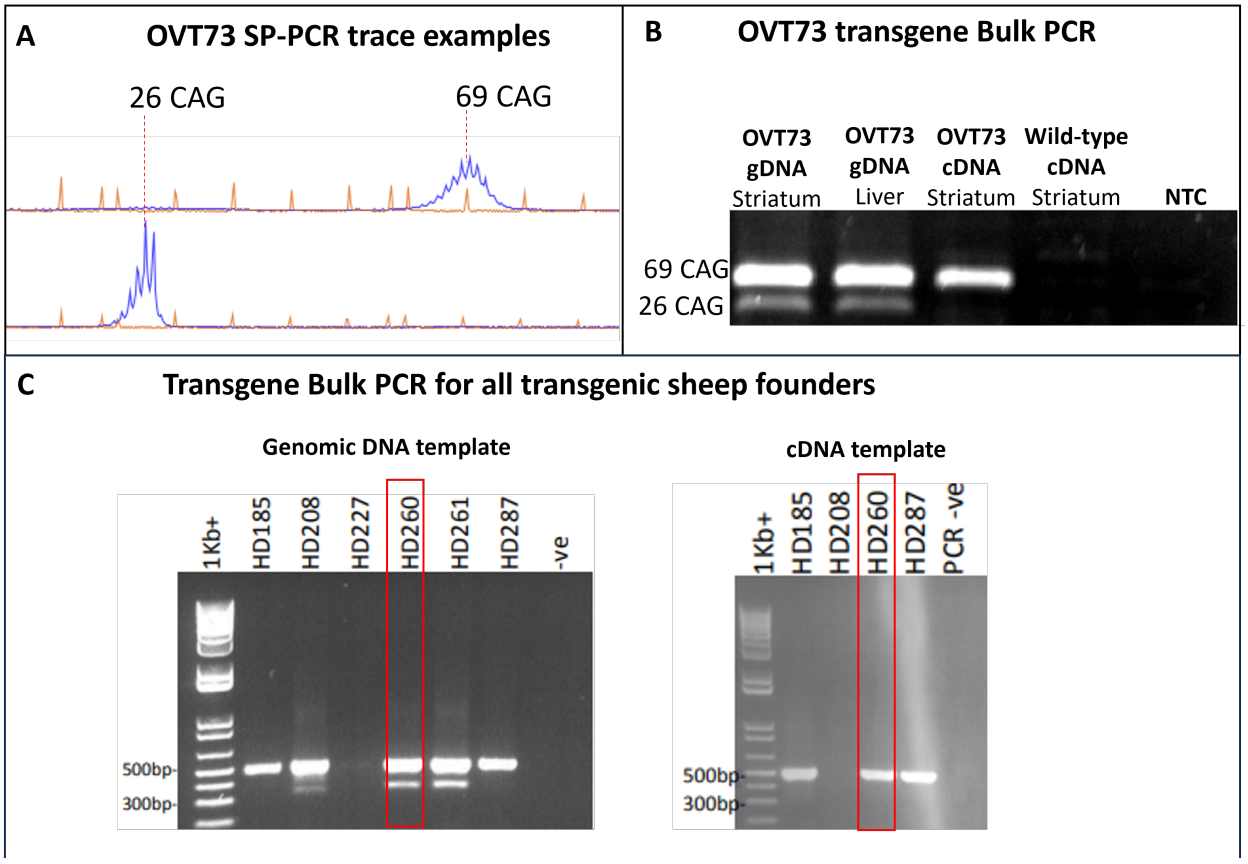
Age (y = years, m = months), Generation (G), Analysis method: PCR = polymerase chain reaction, SP = small pool.

Supplementary Table 2. Protein Homology and gene expression of key DNA repair genes in sheep (public database query).

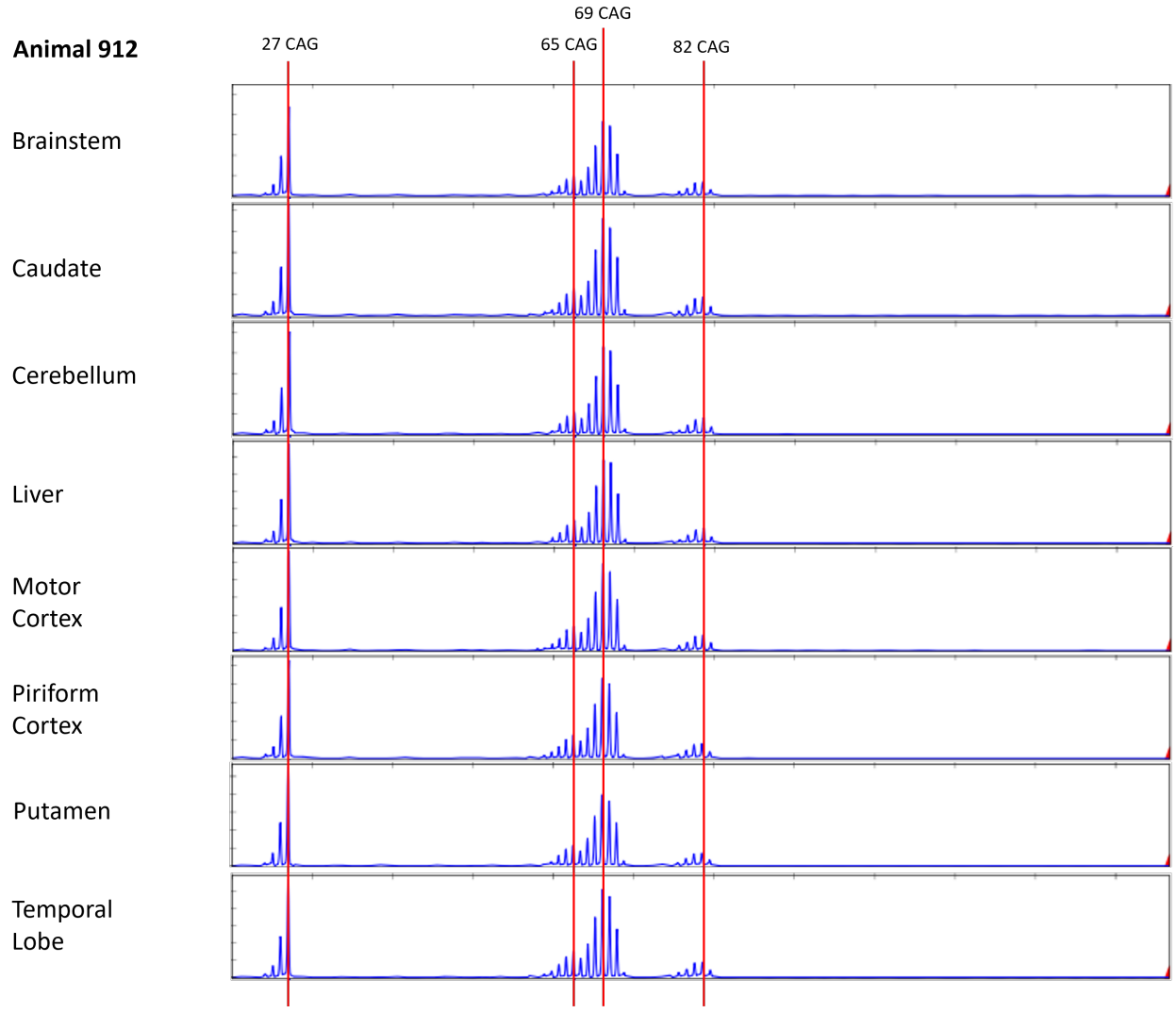
Protein	Human Accession (NCBI)	Sheep Accession (NCBI)	Human – sheep identity	BioGPS url
FAN1	NP_055782.3	XP_042091087.1	71.78%	http://biogps.org/sheepatlas/#goto=genereport&id=101118223
MSH3	NP_002430.3	XP_042106656.1	85.97%	http://biogps.org/sheepatlas/#goto=genereport&id=101121679
MLH1	NP_000240.1	XP_004018267.2	90.63%	http://biogps.org/sheepatlas/#goto=genereport&id=101108344
LIG1	NP_000225.1	XP_027834353.2	83.91%	http://biogps.org/sheepatlas/#goto=genereport&id=101115595
PMS1	NP_000525.1	XP_004004549.2	83.60%	http://biogps.org/sheepatlas/#goto=genereport&id=101108344
PMS2	NP_000526.2	XP_011960028.2	80.23%	http://biogps.org/sheepatlas/#goto=genereport&id=101115595

Human - sheep species homology (homo sapiens, ovis aries) of reference protein sequences for key DNA repair genes influencing CAG repeat instability are shown. Homology (% identity) was determined by Protein BLAST alignment of reference sequences https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome (8 May 2023). Reference protein accession numbers were retrieved from https://www.ncbi.nlm.nih.gov/ (8 May 2023). Gene expression was confirmed by query of each gene name using the search box in the publicly available BioGPS sheep atlas database http://biogps.org/sheepatlas/#goto=welcome. The URL showing expression results for each gene is provided. Gene expression of all genes is readily detectable throughout the body of domestic sheep in this database.

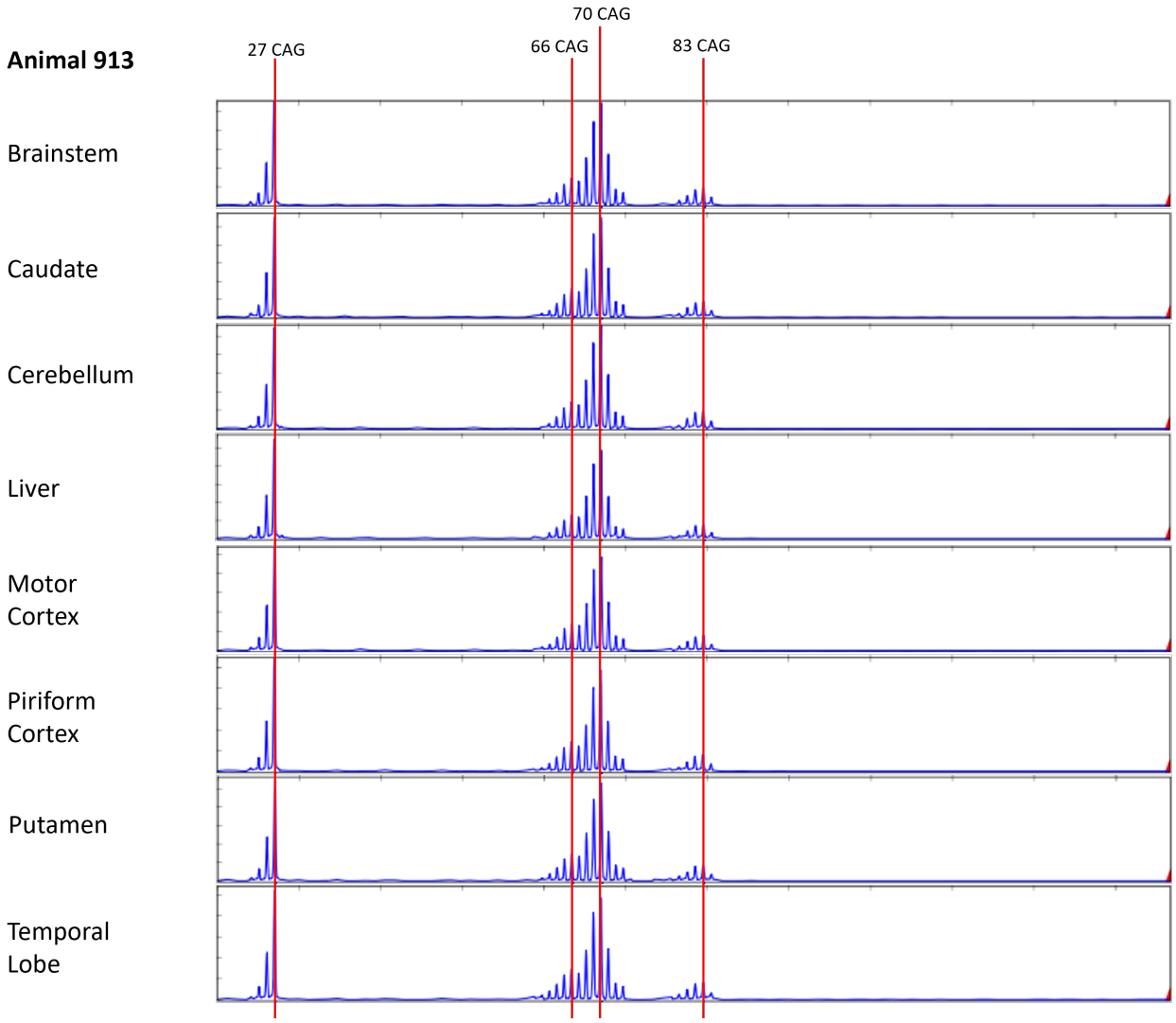
Supplementary Figure 1. A short CAG repeat allele exists at the OVT73 transgene locus but is not expressed. PCR of the OVT73 transgene revealed two dominant transgene alleles with pure CAG repeat tracts of 69 and 26 repeats. A) Examples of GeneScan traces for single molecules from liver samples demonstrate the alternative transgene copies detected by SP-PCR in 5-year-old OVT73 sheep (G1/G2). B) Bulk PCR and RT-PCR gel electrophoresis images from representative OVT73 sheep samples show that both long (69 pure CAG repeats, ~470 bp product) and short (26 pure CAG repeat, ~350 bp product) alleles are detected in genomic DNA, but only the longer allele is detected in complementary DNA (cDNA) derived from total RNA. Lane 1: Genomic DNA OVT73 striatum, Lane 2: Genomic DNA OVT73 liver, Lane 3: cDNA OVT73 striatum, Lane 4: cDNA wild-type sheep striatum, Lane 5: no template PCR control. C) Gel electrophoresis image for bulk genomic DNA PCR (left) and RT-PCR (right) of OVT73 transgene in tail tissue from the founder for this line (HD260, shown in red box) as well as five other founders which were not developed further. Two bands can be seen for genomic DNA in HD260, representing the 26 and 69 CAG alleles. The short allele is present in the three founders with high copy numbers of the 69 CAG allele (HD208, HD260, HD261), and not the founders with lower copy numbers of the transgene (HD185, HD227, HD287). Only the 69 CAG allele is observed in cDNA from HD260 (right), indicating that the short allele is not expressed. –ve lanes are no template PCR controls. 1kb Plus DNA ladder (Invitrogen).



Supplementary Figure 2. Alternative copies of the OVT73 transgene have different polyglutamine coding repeat lengths (Animal 912). Bulk PCR traces from striatal genomic DNA of three 10-year-old (G3) OVT73 sheep indicate presence of multiple copies of the OVT73 transgene with differing polyglutamine-coding repeat lengths, inserted at the ovine chromosome 10 locus. Bulk PCR traces from brain and liver tissues are shown for a single animal (animal 912). The dominant pure CAG repeat length was 69. Repeat lengths of 28 (short allele), 65 and 82 are also observed. Traces were consistent across the eight tissues examined. GeneScan 500 LIZ internal size standard was used to determine product size. CAG repeat size was estimated against knock-in mouse model standards, with adjustment for known differences in the polyglutamine-polyproline repeat sequence structure between the mouse and sheep models (refer to the Supplementary Material above).



Supplementary Figure 3. Alternative copies of the OVT73 transgene have different polyglutamine coding repeat lengths (Animal 913). Bulk PCR traces from striatal genomic DNA of three 10-year-old (G3) OVT73 sheep indicate presence of multiple copies of the OVT73 transgene with differing polyglutamine-coding repeat lengths, inserted at the ovine chromosome 10 locus. Bulk PCR traces from brain and liver tissues are shown for a single animal (animal 913). The dominant pure CAG repeat length was 70. Repeat lengths of 28 (short allele), 66 and 82 are also observed. Traces were consistent across the eight tissues examined. GeneScan 500 LIZ internal size standard was used to determine product size. CAG repeat size was estimated against knock-in mouse model standards, with adjustment for known differences in the polyglutamine-polyproline repeat sequence structure between the mouse and sheep models (refer to the Supplementary Material above).



Supplementary Figure 4. MiSeq analysis of OVT73 transgene polyglutamine-polyproline repeat structure. Miseq sequencing was performed on transgene bulk PCR products from two 10-year-old OVT73 animals (liver genomic DNA template), HD909 and HD913. Frequency of sequence structures determined for each animal is summarized for two analysis approaches. A) Most frequent structures detected (Read Count >10). These reads met stringent QC criteria of including a CAGCAGCAG 9-mer at the 5' end and a CAGCTTCCT 9-mer at the 3' end of the polyglutamine- and polyproline-encoding sequence, on both Forward and Reverse reads (Methodology described in [4]. Highlighted are the most frequent reads for the two most prominent repeat lengths in the sheep. The short transgene has been preferentially amplified. Note that the most frequent short transgene copy for each animal contains 27 CAGs. B) To maximize the detection of long CAG-containing reads, we analyzed Forward reads only and without constraint of the 3' flank as above. This provides evidence for longer uninterrupted repeats. Examples of the most frequent reads with CAGs in the high 70s- 80s are shown. The presence of the specific CAG repeat lengths aligns with bulk PCR fragment analysis and SP-PCR for these animals.

A

HD909 (liver) transgene structure	Read Count
CAG27 CAA1 CAG1 CAA1 CAG1 CCG9 CCT3 CAGCTTCCT1	582
CAG26 CAA1 CAG1 CAA1 CAG1 CCG9 CCT3 CAGCTTCCT1	229
CAG67 CAA1 CAG1 CAA1 CAG1 CCG9 CCT3 CAGCTTCCT1	43
CAG69 CAA1 CAG1 CAA1 CAG1 CCG9 CCT3 CAGCTTCCT1	36
CAG68 CAA1 CAG1 CAA1 CAG1 CCG9 CCT3 CAGCTTCCT1	36
CAG64 CAA1 CAG1 CAA1 CAG1 CCG9 CCT3 CAGCTTCCT1	35
CAG63 CAA1 CAG1 CAA1 CAG1 CCG9 CCT3 CAGCTTCCT1	34
CAG27 CAA1 CAG1 CAA1 CAG1 CCG8 CCT3 CAGCTTCCT1	30
CAG66 CAA1 CAG1 CAA1 CAG1 CCG9 CCT3 CAGCTTCCT1	29
CAG25 CAA1 CAG1 CAA1 CAG1 CCG9 CCT3 CAGCTTCCT1	27
CAG65 CAA1 CAG1 CAA1 CAG1 CCG9 CCT3 CAGCTTCCT1	23
CAG62 CAA1 CAG1 CAA1 CAG1 CCG9 CCT3 CAGCTTCCT1	18
CAG26 CAA1 CAG1 CAA1 CAG1 CCG8 CCT3 CAGCTTCCT1	15
CAG24 CAA1 CAG1 CAA1 CAG1 CCG9 CCT3 CAGCTTCCT1	14
CAG28 CAA1 CAG1 CAA1 CAG1 CCG9 CCT3 CAGCTTCCT1	13
CAG70 CAA1 CAG1 CAA1 CAG1 CCG9 CCT3 CAGCTTCCT1	13
CAG61 CAA1 CAG1 CAA1 CAG1 CCG9 CCT3 CAGCTTCCT1	12

HD913 (liver) transgene structure	Read Count
CAG27 CAA1 CAG1 CAA1 CAG1 CCG9 CCT3 CAGCTTCCT1	808
CAG26 CAA1 CAG1 CAA1 CAG1 CCG9 CCT3 CAGCTTCCT1	303
CAG25 CAA1 CAG1 CAA1 CAG1 CCG9 CCT3 CAGCTTCCT1	74
CAG69 CAA1 CAG1 CAA1 CAG1 CCG9 CCT3 CAGCTTCCT1	65
CAG68 CAA1 CAG1 CAA1 CAG1 CCG9 CCT3 CAGCTTCCT1	57
CAG27 CAA1 CAG1 CAA1 CAG1 CCG8 CCT3 CAGCTTCCT1	52
CAG67 CAA1 CAG1 CAA1 CAG1 CCG9 CCT3 CAGCTTCCT1	33
CAG65 CAA1 CAG1 CAA1 CAG1 CCG9 CCT3 CAGCTTCCT1	29
CAG64 CAA1 CAG1 CAA1 CAG1 CCG9 CCT3 CAGCTTCCT1	27
CAG63 CAA1 CAG1 CAA1 CAG1 CCG9 CCT3 CAGCTTCCT1	24
CAG24 CAA1 CAG1 CAA1 CAG1 CCG9 CCT3 CAGCTTCCT1	22
CAG70 CAA1 CAG1 CAA1 CAG1 CCG9 CCT3 CAGCTTCCT1	21
CAG66 CAA1 CAG1 CAA1 CAG1 CCG9 CCT3 CAGCTTCCT1	17
CAG28 CAA1 CAG1 CAA1 CAG1 CCG9 CCT3 CAGCTTCCT1	13
CAG26 CAA1 CAG1 CAA1 CAG1 CCG8 CCT3 CAGCTTCCT1	12

B

HD909 (liver) transgene structure	Read Count
76CAG 1CAA 1CAG 1CAA 1CAG 9CCG	41
80CAG 1CAA 1CAG 1CAA 1CAG	33
75CAG 1CAA 1CAG 1CAA 1CAG 9CCG	32
82CAG 1CAA 1CAG 1CAA 1CAG	29
81CAG 1CAA 1CAG 1CAA 1CAG	28

HD913 (liver) transgene Structure	Read Count
82CAG 1CAA 1CAG 1CAA 1CAG	107
81CAG 1CAA 1CAG 1CAA 1CAG	85
80CAG 1CAA 1CAG 1CAA 1CAG	69
83CAG 1CAA 1CAG 1CAA 1CAG	41
79CAG 1CAA 1CAG 1CAA 1CAG	26

Supplementary Dataset 1. Frequency table of CAG repeat lengths detected in single DNA molecules containing the OVT73 transgene by small-pool PCR (5-year-old sheep liver and striatum). See Excel file.