

Supplementary file 1

Estimation of the percentage of on-target and full-length reads for each experiment

To estimate the percentage of on-target and full-length reads for each experiment (*i.e.* a particular number of CAGs sequenced on a particular sequencing platform) we applied the following steps on sequencing reads obtained from one representative sample per experiment (the representative sample analysed was the cerebellum of the older mice):

Sequencing reads were first aligned (using BWA-MEM with default parameters as for all alignments mentioned below) to a reference sequence containing the 5'-flank and 1000 CAGs (Table SF1.1). Reads aligned to the 5'-flank were considered on-target and then aligned to a reference sequence containing the 3'-flank and 1000 CAGs (Table SF1.1).

Reads aligned to both flanks (determined based on the left-most alignment position) were considered full-length. Reads not aligned to the 5'-flank were aligned to the reference sequence containing the 3'-flank and 1000 CAGs (Table SF1.1). Reads aligned to at least one flank were considered on-target but not full-length. Reads not aligned to either flanks were then aligned to a pure CAG repeat reference sequence (Table SF1.1). Reads aligned to pure CAG repeat were considered on-target.

Table SF1.1: Reference sequence used

>5'-flank and 1000 CAGs GCCCAGAGCCCCATTCATTGCCCCGGTGCTGAGCGGCGCCGCGAGTCGGCCCCGAGGCCTCCGGGGA CTGCCGTGCCGGGCGGGAGACCGCCATGGCGACCCTGGAAAAGCTGATGAAGGCCTTCGAGTCCCT CAAGTCCTTC (CAG) ₁₀₀₀
>3'-flank and 1000 CAGs GCCATCCCCGCCGTAGCCTGGGACCCGCCGGGACAGGGAGCTGCAGCGGGCCCAAACCTCACGGTTCG GTGCAGCGGCTCCTCAGCCACAGCCGGGCCGGGTGGCGGGCGGGGCGGCGGCGGGGGCGGCTGCGG CTGAGGCAGCAGCGGCTGTGCCTGCGGCGGGCGGCTGAGGAAGCTGAGGAGGCGGCGGGCGGCGGCGG CGGCGGTGGCGGCTGTTG (CTG) ₁₀₀₀
>pure CAG repeat (CAG) ₁₀₀₀

Note: The 3'-flank and 1000 CAGs was considered in reverse complement orientation so that the 3'-flank would be on the left which then allowed to filter reads based on the POS column of the SAM alignment files which corresponds to the leftmost alignment position for a reads relatively to the reference sequence used.

Reads that did not align to either flanks, or to the pure CAG repeat reference sequence, were considered off-target. Assuming that the most likely source of off-target reads would be non-specific PCR of mouse DNA, the most likely source of the off-target reads was determined using Blastn {Johnson, 2008 #199} against all *Mus musculus* sequences available in NCBI Nucleotide collection (nr/nt) {Sayers, 2020 #198}. If no *Mus musculus* match was obtained, we attempted to determine the most likely source of the off-target reads using Blastn {Johnson, 2008 #199} against all sequences available in NCBI Nucleotide collection (nr/nt) {Sayers, 2020 #198}.

Result obtained are presented in Table SF1.2 below.

Table SF1.2: Number and percentage of on-target and full-length reads as well as source of the off-target reads for each experiment

Experiment	Number of reads generated per sample	Number of reads on-target	Number of full-length reads	Number of reads off-target	Source of the off-target reads
~ 55 CAG repeats on MiSeq	112,113	111,873 (99.78%)	111,573 (99.52%)	240 (0.21%)	210 (0.19%) were primer-dimers and 30 (0.02%) remained unidentified
~ 55 CAG repeats on PacBio	728	728 (100%)	720 (98.80%)	0	NA
~ 110 CAG repeats on MiSeq	5262	5262 (100%)	529 (10.05%)	0	NA
~ 110 CAG repeats on PacBio	1716	1715 (99.94%)	1,561 (90.97%)	1 (0.06%)	<i>Mus musculus Foxe1</i> gene for forkhead box E1
~ 255 CAG repeats on PacBio	812	811 (99.88%)	541 (66.63%)	1 (0.12%)	<i>Mus musculus Foxe1</i> gene for forkhead box E1
~ 470 CAG repeats on PacBio	580	577 (99.48%)	419 (72.24%)	3 (0.52%)	<i>Mus musculus Foxe1</i> gene for forkhead box E1

References

- Johnson, M., I. Zaretskaya, Y. Raytselis, Y. Merezhuk, S. McGinnis *et al.*, 2008 NCBI BLAST: a better web interface. *Nucleic acids research* 36: W5-9.
- Sayers, E. W., J. Beck, J. R. Brister, E. E. Bolton, K. Canese *et al.*, 2020 Database resources of the National Center for Biotechnology Information. *Nucleic acids research* 48: D9-d16.