Supplementary Materials and Methods

Quantitation of mRNA

Mouse organs were homogenized using Precellys lysing kits (Bertin Technologies, Rockville, MD). Total RNA was isolated using Maxwell®16 LEV simplyRNA tissue Kit (Promega) as per manufacturer's instructions. The RNA quality and concentration was determined using a Nanodrop Spectrophotometer (DeNovix). The total RNA (400 ng) was reverse transcribed in 20 µL final volume using SuperScript®VILO™ cDNA synthesis kit (Thermo Fisher Scientific) following the manufacturer's protocol. Real time PCR was done in triplicate using TaqMan® Fast Advanced Master mix (Thermo Fisher Scientific), 2 µL of the cDNA, the Taqman probe-primer pairs (Thermo Fisher Scientific) of the *Gapdh* (Mouse GAPDH Endogenous Control, VIC®/MGB Probe, primer limited), and the Taqman probe-primer pairs either for *Fan1* (Mm00625959_m1) or *Exo1* (Mm00516302_m1) on StepOnePlus Real-Time PCR System (Thermo Fisher Scientific) following the manufacturer's protocol. For comparison of *Fan1* mRNA expression in brain of $Fan1^{D/D}$ and $Fan1^{A/A}$ mice, relative *Fan1* mRNA levels were normalized to the endogenous *Gapdh* level in the brain of different animals. Statistical analyses of the differences in mRNA level in $Fan1^{D/D}$ and $Fan1^{A/A}$ mice were performed using GraphPad Prism 8.4. Statistical significance was assessed using the two-tailed unpaired t test. For comparison of mRNA expression in different tissue, equal amounts of RNA were used for each determination. Normalization of *Fan1* mRNA expression level in different regions of mouse brain was carried out by comparing the Ct value in striatum, cortex, cerebellum to the Ct value of *Fan1* obtained from total brain. Normalization of *Fan1* and *Exo1* expression levels in different organs was carried out by comparing the Ct value for the mRNA in different organs to the Ct value of *Exo1*

obtained from brain. *Gapdh* was used as endogenous control to compare the expression level of *Fan1* and *Exo1* in the same tissue.

Supplementary Figure Legends

Fig. S1. **Generation of the FAN1^{A963} mouse.** A) Comparison of sequences of the region of *Fan1* that encodes D963 and the A963 mutant. The residues shown in red correspond to the mutations introduced into the *Fan1* allele in the D963A mutant mice. B) Sequence traces from this region in *Fan1*$^{D/D}$, *Fan1*$^{D/A}$ and *Fan1*$^{A/A}$ mice. C) Location of primers used to genotype the mice. The AF and DR primers overlap the modified region and are specific for the A and D alleles respectively. D) Agarose gel electrophoresis of the PCR amplification products from genomic DNA of mice carrying D/D, D/A of A/A alleles with different combinations of primers specific for the WT and mutant alleles as indicated.

Fig. S2. **CRISPR editing does not disrupt proper splicing of the mutated exon.** A) Sequence alignment of the WT *Fan1* DNA sequence with the cDNA produced from the Fan1 transcript from *Fan1*$^{A/A}$ mice in the region spanning the exon 12-exon 13 junction and exon 13-exon 14 junction. Exon 13 is underlined. B) Sequence traces from this region in *Fan1*$^{A/A}$ mice.

Fig. S3. **The nuclease domain of FAN1 is evolutionarily conserved.** A) Organization of the FAN1 protein in mice showing the location of the D963 residue that was mutated. The

arrangement of the ubiquitin binding (UBZ), SAF-A/B, Acinus and PIAS (SAP), the tetratricopeptide repeat (TPR) and the virus-type replication-repair nuclease (VRR_NUC) domains is conserved in complex eukaryotes. The PDXn(D/E)XK nuclease motif containing residue D963 is embedded in the VRR_NUC domain. B) Comparison of the C-terminal end of the FAN1 protein containing the PDXn(D/E)XK nuclease motif in *Mus musculus* (mouse), *Homo sapiens* (human), *Danio rerio* (zebrafish), *Caenorhabditis elegans* (roundworm) and the bacterium *Pseudomonas aeruginosa*. The amino acids shown in red are the invariant bases within the PDXn(D/E)XK motif. An asterisk indicates positions that have a residue conserved in all 5 species. A colon indicates conservation between groups of strongly similar properties - roughly equivalent to scoring > 0.5 in the Gonnet PAM 250 matrix. A period indicates conservation between groups of weakly similar properties - roughly equivalent to scoring =< 0.5 and > 0 in the Gonnet PAM 250 matrix. C) Comparison of the C-terminal 2/3 of the mouse and human FAN1 proteins showing 84% sequence identity and 91% sequence similarity across the whole region that includes the SAP domain, the TPR domain and the VRR-NUC domain in which the PDXn(D/E)XK nuclease motif is embedded. The red arrowhead indicates the D963 residue mutated in the FAN1^{A963} mice.

Fig. S4. The D963A mutation in FAN1 significantly increases expansions in the FXD mouse model. Representative repeat PCR profiles from different organs of *Fan1*$^{D/D}$, *Fan1*$^{D/A}$, *Fan1*$^{A/A}$ and *Fan1*$^{-/-}$ males at 6-month of age. The dotted line represents the size of the

original inherited allele as ascertained from the tail DNA taken at 3 weeks of age. The numbers alongside the 6-month old samples indicates the number of repeats added to alleles in that organ during the lifetime of the mouse. The original allele sizes are 157, 161, 159, and 158 for the *Fan1*$^{D/D}$, *Fan1*$^{D/A}$, *Fan1*$^{A/A}$ and *Fan1*$^{-/-}$ males respectively.

Fig. S5. ***Fan1*$^{-/-}$ mice have more extensive expansions in liver than D963A mice.** Pairwise comparison of the PCR profiles from livers of 5 pairs of 6-month-old mice *Fan1* $^{A/A}$ and *Fan1*$^{-/-}$ mice matched exactly for inherited repeat number. The red dotted line represents the size of the original inherited allele as ascertained from the tail DNA taken at 3 weeks of age. The black dotted line indicates the modal size of the expanded allele in *Fan1*$^{A/A}$ mice.

Fig. S6. **The stability of the *Fan1* transcript is unaffected by the mutations introduced.** The amount of *Fan1* mRNA in brain of 6-month old male mice was evaluated by real-time quantitative PCR as described in the Materials and Methods. The *Fan1* mRNA levels are expressed relative to the levels of a house keeping gene, *Gapdh*. The mRNA levels are an average of the levels from 5 *Fan1*$^{D/D}$, and 5 *Fan1*$^{A/A}$ mice. The error bars indicate the standard deviations of the mean. Significance was assessed using two-tailed unpaired *t* tests as described in the Materials and Methods. The difference in *Fan1* mRNA expression between *Fan1*$^{D/D}$ and *Fan1*$^{A/A}$ animals was not significant (*P*=0.5412).

Fig. S7. The levels of *Fan1* mRNA are similar in different brain regions of FXD mice. Relative *Fan1* mRNA levels in different brain regions of 6-month old male mice. The amount of *Fan1* mRNA was evaluated by real-time quantitative PCR as a function of total RNA and the values expressed relative to the levels in total brain. The mRNA levels are an average of the levels from 4 male mice. Error bars represent the standard deviations of the mean.

Fig. S8. Heterozygosity for the D963A allele results in a significant increase in the extent of expansion in different brain regions. A) Representative repeat PCR profiles from different brain regions of *Fan1*$^{D/D}$, *Fan1*$^{D/A}$ and *Fan1*$^{A/A}$ males at 6-month of age. The dotted line represents the size of the original inherited allele as ascertained from the tail DNA taken at 3 weeks of age. The numbers alongside the samples from the animals at 6-months of age indicates the number of repeats added during the lifetime of the mouse. The original allele sizes are 163, 161, and 162 for the *Fan1*$^{D/D}$, *Fan1*$^{D/A}$ and *Fan1*$^{A/A}$ males respectively. B-C) Comparison of the number of repeats added (B) and Expansion Index (C) in different brain regions of 6-month old *Fan1*$^{D/D}$, *Fan1*$^{D/A}$ and *Fan1*$^{A/A}$ mice with ~161 repeats in the original allele. The data represent the average of data from 3 animals of each genotype with 160-164 repeats. The error bars indicate the standard deviations of the mean. Significance was assessed using a RM two-way ANOVA (genotype and tissue as variables) with correction for multiple testing as described in the Materials and Methods. The adjusted p values for the genotype effect are listed in the table.

Fig. S9. No effect of the loss of FAN1 or EXO1 is seen in an *Msh3* null background. Representative repeat PCR profiles from different organs of WT, *Fan1*$^{-/-}$, *Exo1*$^{-/-}$ or *Msh3*$^{-/-}$ as well as *Msh3*$^{-/-}$ *Fan1*$^{-/-}$and *Msh3*$^{-/-}$ *Exo1*$^{-/-}$ double knockout mice at 6-month of age. The dotted line represents the size of the original inherited allele as ascertained from the tail DNA taken at 3 weeks of age. The numbers alongside the samples from the animals at 6-months of age indicates the number of repeats added during the lifetime of the mouse. The original allele sizes are 168, 168, 169, 168, 168, and 168 for the WT, *Fan1*$^{-/-}$, *Exo1*$^{-/-}$ or *Msh3*$^{-/-}$ as well as *Msh3*$^{-/-}$ *Fan1*$^{-/-}$and *Msh3*$^{-/-}$ *Exo1*$^{-/-}$ double knockout mice respectively.

Fig. S10. Loss of both FAN1 and EXO1 result in more extensive expansions in small intestine than either single mutant alone. Representative repeat PCR profiles from different organs of WT, *Fan1*$^{-/-}$, *Exo1*$^{-/-}$ and double mutant *Fan1*$^{-/-}$ *Exo1*$^{-/-}$ mice at 6-month of age. The dotted line represents the size of the original inherited allele as ascertained from the tail DNA taken at weaning (3-wk tail). The numbers alongside the samples from the animals at 6-months of age indicates the number of repeats added to alleles in the indicated organ during the lifetime of the mouse. The original allele sizes are 163, 165, 164, 165 for the WT, *Fan1*$^{-/-}$, *Exo1*$^{-/-}$ and double mutant *Fan1*$^{-/-}$ *Exo1*$^{-/-}$, mice respectively.

Fig. S11. Comparison of *Fan1* and *Exo1* mRNA levels in different mouse tissues. The amount of *Fan1* mRNA and *Exo1* mRNA in different tissues of 6-month old male mice was evaluated by real-time quantitative PCR as a function of total RNA and the values expressed relative to the levels of *Exo1* transcript in brain. The mRNA levels are an average of the levels from 5 male mice. Error bars represent the standard deviation of the mean. The boxes above the black bars indicate the fold difference in mRNA level in the indicated tissue of *Fan1* mRNA relative to the same tissue of *Exo1* mRNA.

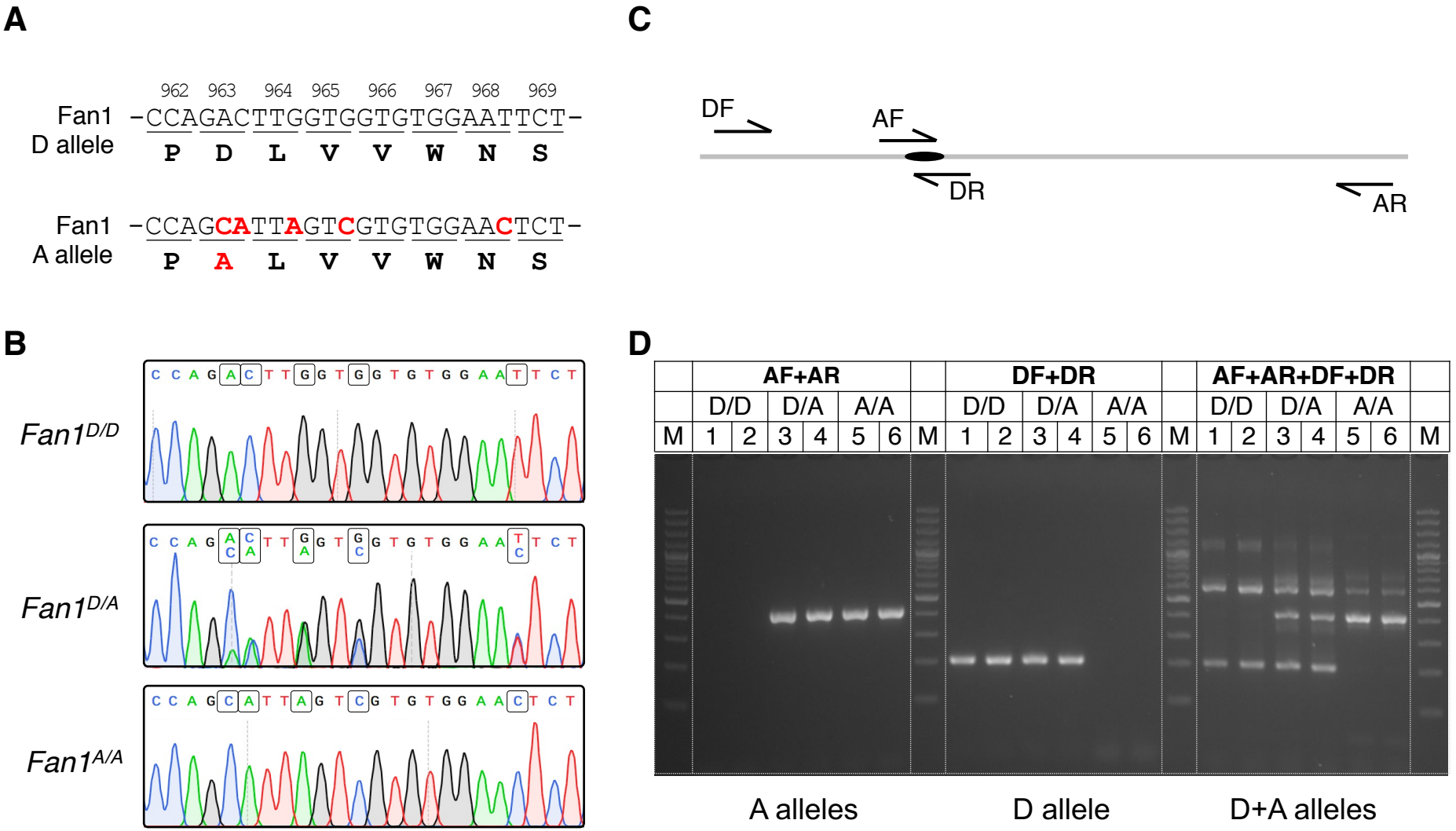

A
962 963 964 965 966 967 968 969
Fan1 -CCAGACTTGGTGGTGTGGAATTCT-
D allele
P D L V V W N S
Fan1 -CCAGCATTAGTCGTGTGGAACTCT-
A allele
P A L V V W N S
B
Fan1D/D
Fan1D/A
Fan1A/A
C
DF
AF
DR
AR
D
AF+AR
DF+DR
AF+AR+DF+DR
D/D D/A A/A
M 1 2 3 4 5 6 M
A alleles
D allele
D+A alleles

Fig.S1

A

```
                                        Exon12 | Exon13
WT ---GACCGCTTCACCTCCCTACAGCAAGCTCAGGATCTTGTCTCCTGCCTCGGGGGTCCTGTCCTCAGTGGTGTGTGCAGGCGCCTGGCTGCTGACTT
mutant ----------------------------------------------------------------------------------------------

                                                                  Exon13 | Exon14
WT TCGGCACTGCCGAGGGGGCCTCCCAGACTTGGTGGTGTGGAATTCTCAGAGCCACCATTGCAAGCTGGTGGAGGTGAAAGGCCCCAGTGATCGA----
mutant ----------------------CA--A--C--------C--------------------------------------------------
```

B

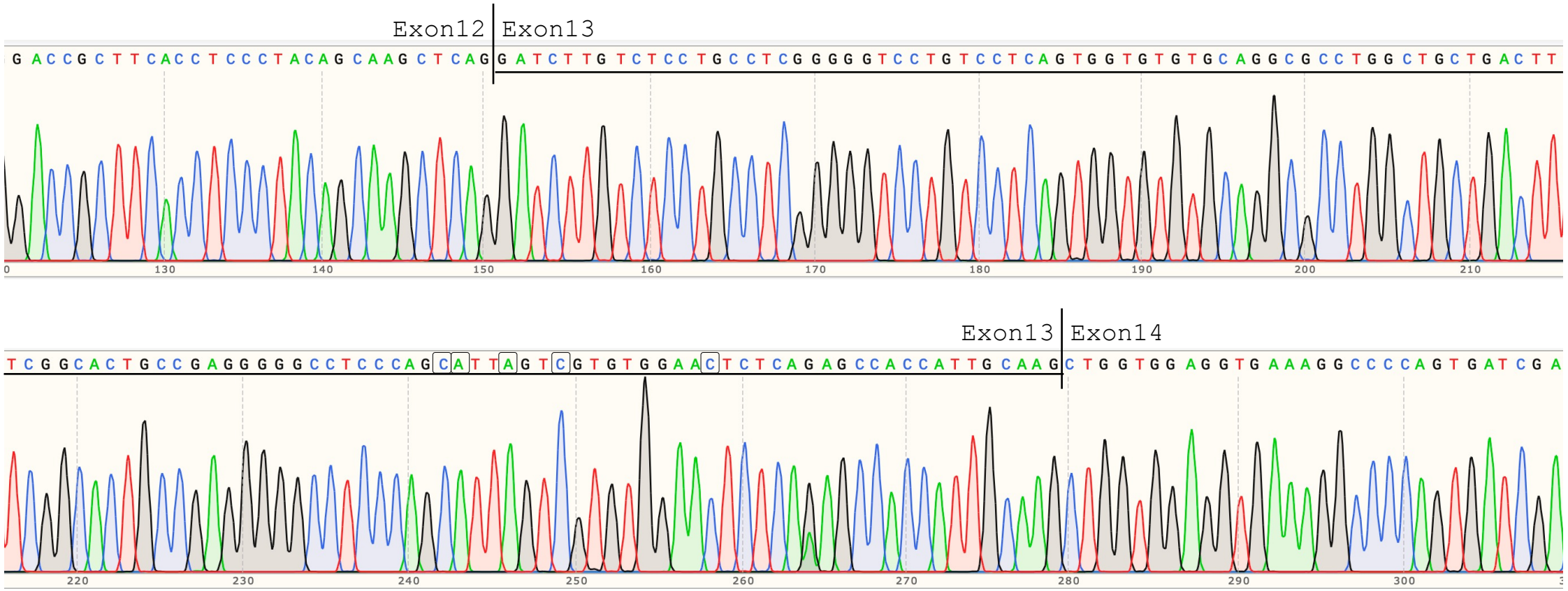
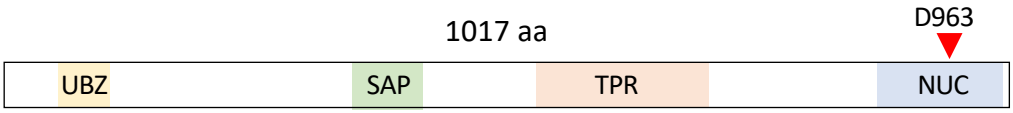


Fig.S2

A



B

```
M. musculus      936 CLGGPVLSGVCRRLAADFRHCRGGLPDLVVWNSQSHHCKLVEVKGPSDRLSCKQMIWLYELQKLGADVEVCHVVA 1012
H. sapiens       933 CLGGPVLSGVCRHLAADFRHCRGGLPDLVVWNSQSRHFKLVEVKGPNDRLSHKQMIWLAELQKLGAEVEVCHVVA 1009
D. rerio         896 CLGGHFLSGVFLRMAKDYRHCRGGLPDLVVWSTSSNKYKLVEVKGPNDRLSQKQQIWLDELRKLGADVEVCHVTA 980
C. elegans       788 CIPRPALILILRRLAENYRNSRSGFPDLTLWNPETKRVAVVEVKGPGDRLSTKQRLWLAIFADSGIRAEVCHVAA 859
P. aeruginosa    477 CLPAAHLRAWFERLLEDIPGNRAGLPDLIQFWPAQRRYRMVEVKGPGDRLQDNQLRWLQFCRREREMPVAVCYVRW 555
                     *:    *     ::  :    *.*:***  :    .:  :******.***. :*  **    .     .**:*
```

C

Score	Expect	Method	Identities	Positives	Gaps
1127 bits(2914)	0.0	Compositional matrix adjust.	545/645(84%)	592/645(91%)	0/645(0%)

```
Mouse  376 PYYLRSFLVVLQALLGNEEDMKLFDEQEKAIITRFYQLSASGQKLYVRLFQRKLTWIKMS 435
           PYYLRSFLVVL+ +L NE+DM LFDEQEK I+T+FYQLSA+GQKLYVRLFQRKL+WIKM+
Human  373 PYYLRSFLVVLKTVLENEDDMLLFDEQEKGIVTKFYQLSATGQKLYVRLFQRKLSWIKMT 432

Mouse  436 KLEYEEIASDLTPVVEELKDSGFLQTESELQELSDVLELLSAPELKALAKTFHLVSPGGQ 495
           KLEYEEIA DLTPV+EEL ++GFLQTESELQELS+VLELLSAPELK+LAKTFHLV+P GQ
Human  433 KLEYEEIALDLTPVIEELTNAGFLQTESELQELSEVLELLSAPELKSLAKTFHLVNPNGQ 492

Mouse  496 KQQLVDAFHKLAKQRSVCTWGKTQPGIRAVILKRAKDLAGRSLRVCKGPRAVFARILLLF 555
           KQQLVDAF KLAKQRSVCTWGK +PGI AVILKRAK LAG+S+R+CKGPRAVF+RILLLF
Human  493 KQQLVDAFLKLAKQRSVCTWGKNKPGIGAVILKRAKALAGQSVRICKGPRAVFSRILLLF 552

Mouse  556 SLTDSMEDEEAACGGQGQLSTVLLVNLGRMEFPQYTICRKTQIFRDREDLIRYAAAAHML 615
           SLTDSMEDE+AACGGQGQLSTVLLVNLGRMEFP YTI RKT IF+DR+DLIRYAAA HML
Human  553 SLTDSMEDEDAACGGQGQLSTVLLVNLGRMEFPSYTINRKTHIFQDRDDLIRYAAATHML 612

Mouse  616 SDISAAMASGNWEDAKELARSAKRDWEQLKSHPSLRYHEALPPFLRCFTVGWIYTRISSR 675
           SDIS+AMA+GNWE+AKELA+ AKRDW +LK+HPSLR HE LP FLRCFTVGWIYTRI SR
Human  613 SDISSAMANGNWEEAKELAQCAKRDWNRLKNHPSLRCHEDLPLFLRCFTVGWIYTRILSR 672

Mouse  676 AVEVLERLHMYEEAVKELENLLSQKIYCPDSRGRWWDRLALNLHQHLKRLEEAIRCIREG 735
            VE+L+RLHMYEEAV+ELE+LLSQ+IYCPDSRGRWWDRLALNLHQHLKRLE  I+CI EG
Human  673 FVEILQRLHMYEEAVRELESLLSQRIYCPDSRGRWWDRLALNLHQHLKRLEPTIKCITEG 732

Mouse  736 LADPHVRTGHRLSLYQRAVRLRESPSCRKYKHLFSRLPEVAVGDVKHVTITGRLCPQHGM 795
           LADP VRTGHRLSLYQRAVRLRESPSC+K+KHLF +LPE+AV DVKHVTITGRLCPQ GM
Human  733 LADPEVRTGHRLSLYQRAVRLRESPSCKKFKHLFQQLPEMAVQDVKHVTITGRLCPQRGM 792

Mouse  796 GKSVFVMESGDGANPTTVLCSVEELALGYYRQSGFDQGIHGEGSTFSTLCGLLLWDIIFM 855
            KSVFVME+G+ A+PTTVLCSVEELAL +YR+SGFDQGIHGEGSTFSTL GLLLWDIIFM
Human  793 CKSVFVMEAGEAADPTTVLCSVEELALAHYRRSGFDQGIHGEGSTFSTLYGLLLWDIIFM 852

Mouse  856 DGIPDVFRNAYQASPLDLLTDSFFASREQALEARLQLIHSAPAESLRAWVGEAWQAQQGR 915
           DGIPDVFRNA QA PLDL TDSFF SR  ALEARLQLIH AP ESLRAWV   W  Q+GR
Human  853 DGIPDVFRNACQAFPLDLCTDSFFTSRRPALEARLQLIHDAPEESLRAWVAATWHEQEGR 912

Mouse  916 VASLVSWDRFTSLQQAQDLVSCLGGPVLSGVCRRLAADFRHCRGGLPDLVVWNSQSHHCK 975
           VASLVSWDRFTSLQQAQDLVSCLGGPVLSGVCR LAADFRHCRGGLPDLVVWNSQS H K
Human  913 VASLVSWDRFTSLQQAQDLVSCLGGPVLSGVCRHLAADFRHCRGGLPDLVVWNSQSRHFK 972
                                                    **
Mouse  976 LVEVKGPSDRLSCKQMIWLYELQKLGADVEVCHVVAVGAKSKGLG  1020
           LVEVKGP+DRLS KQMIWL ELQKLGA+VEVCHVVAVGAKS+ L
Human  973 LVEVKGPNDRLSHKQMIWLAELQKLGAEVEVCHVVAVGAKSQSLS  1017
             * *
```

SAP | TPR | VRR_NUC | PDXn(D/E)XK motif

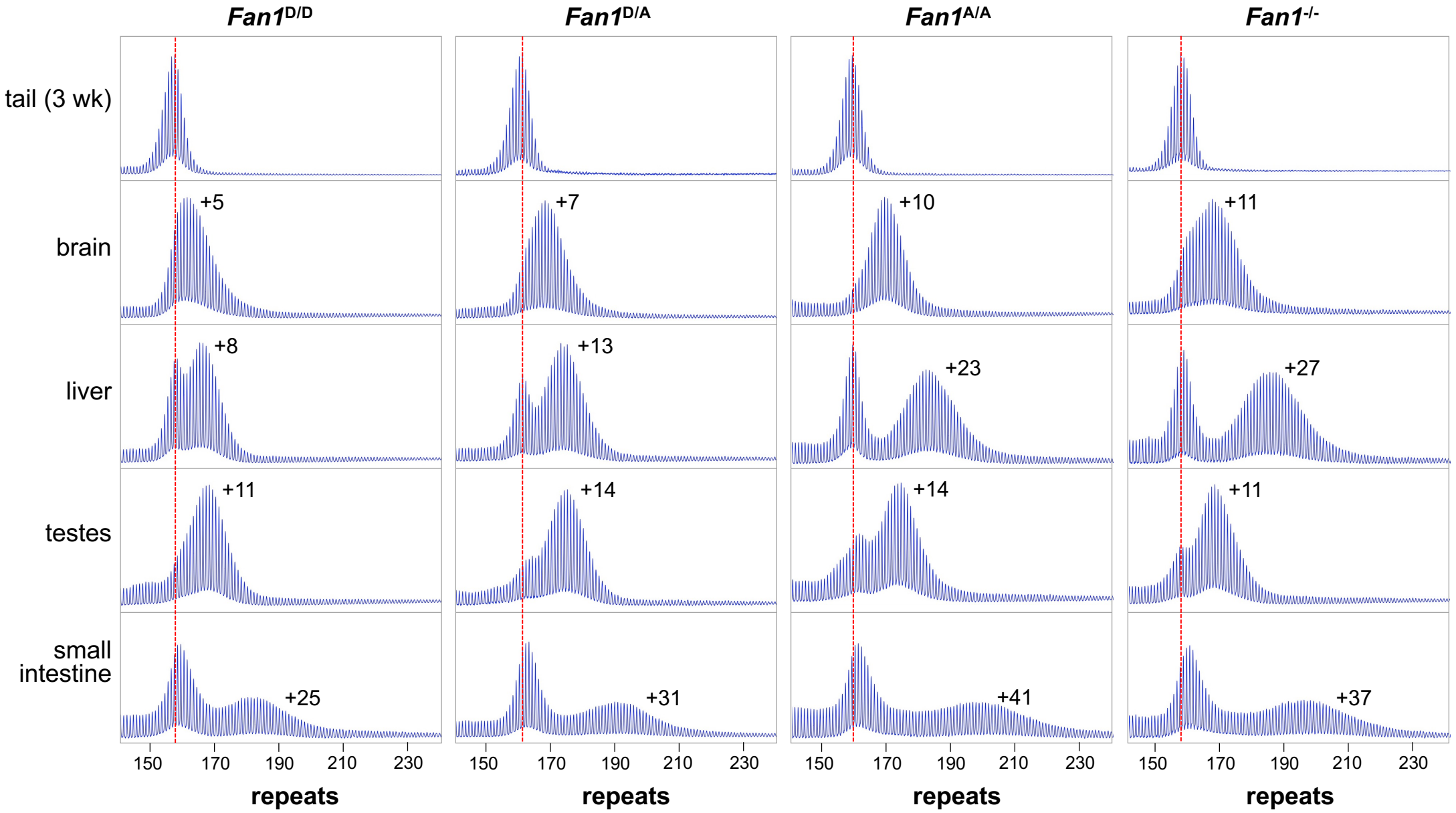

Fan1D/D
Fan1D/A
Fan1A/A
Fan1-/-
tail (3 wk)
brain
liver
testes
small intestine
+5
+7
+10
+11
+8
+13
+23
+27
+11
+14
+14
+11
+25
+31
+41
+37
150
170
190
210
230
repeats

Fig. S4

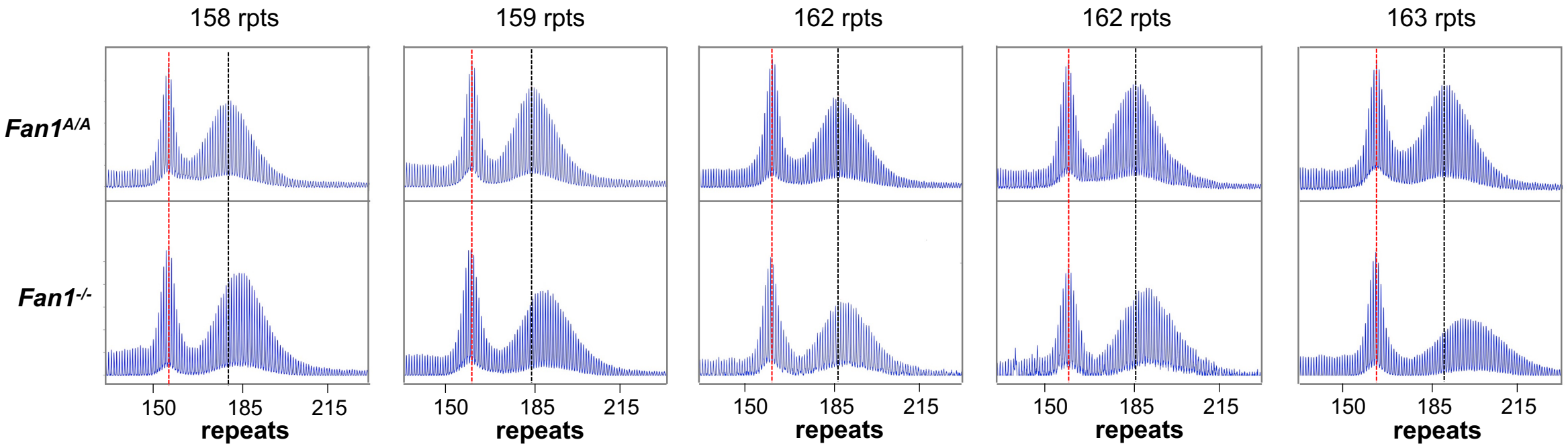

158 rpts
159 rpts
162 rpts
162 rpts
163 rpts
Fan1A/A
Fan1-/-
150
185
215
repeats

Fig.S5

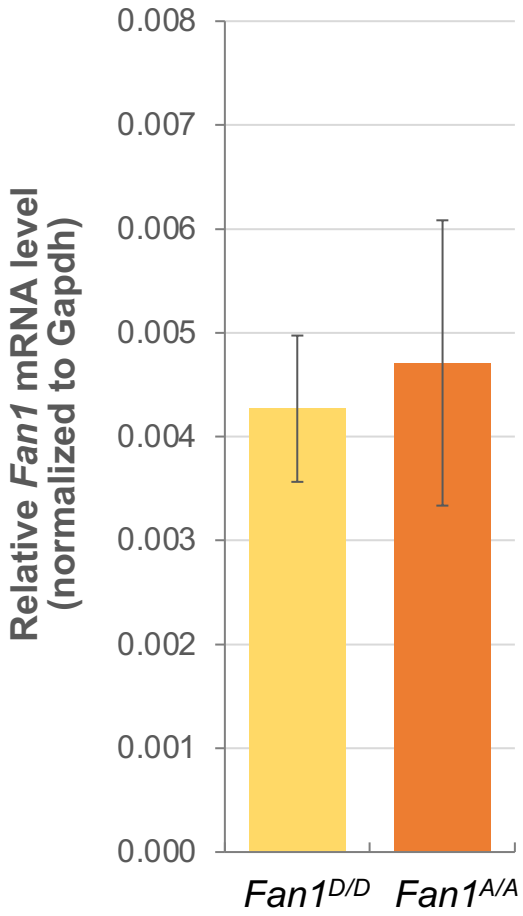

Relative Fan1 mRNA level
(normalized to Gapdh)
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0.000
Fan1D/D
Fan1A/A

Fig.S6

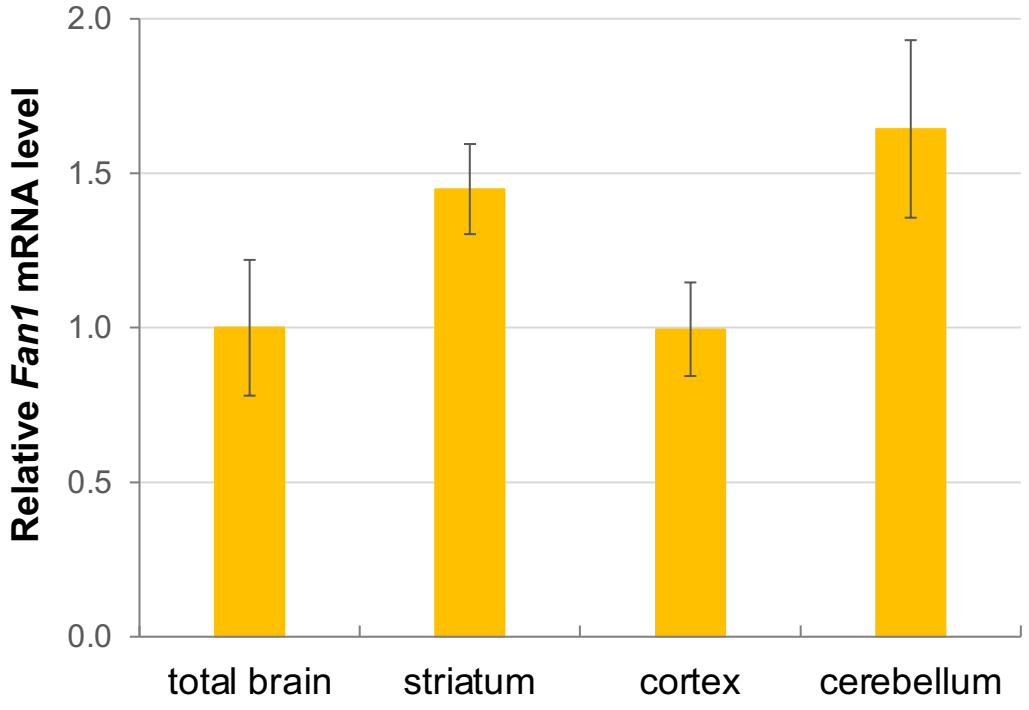
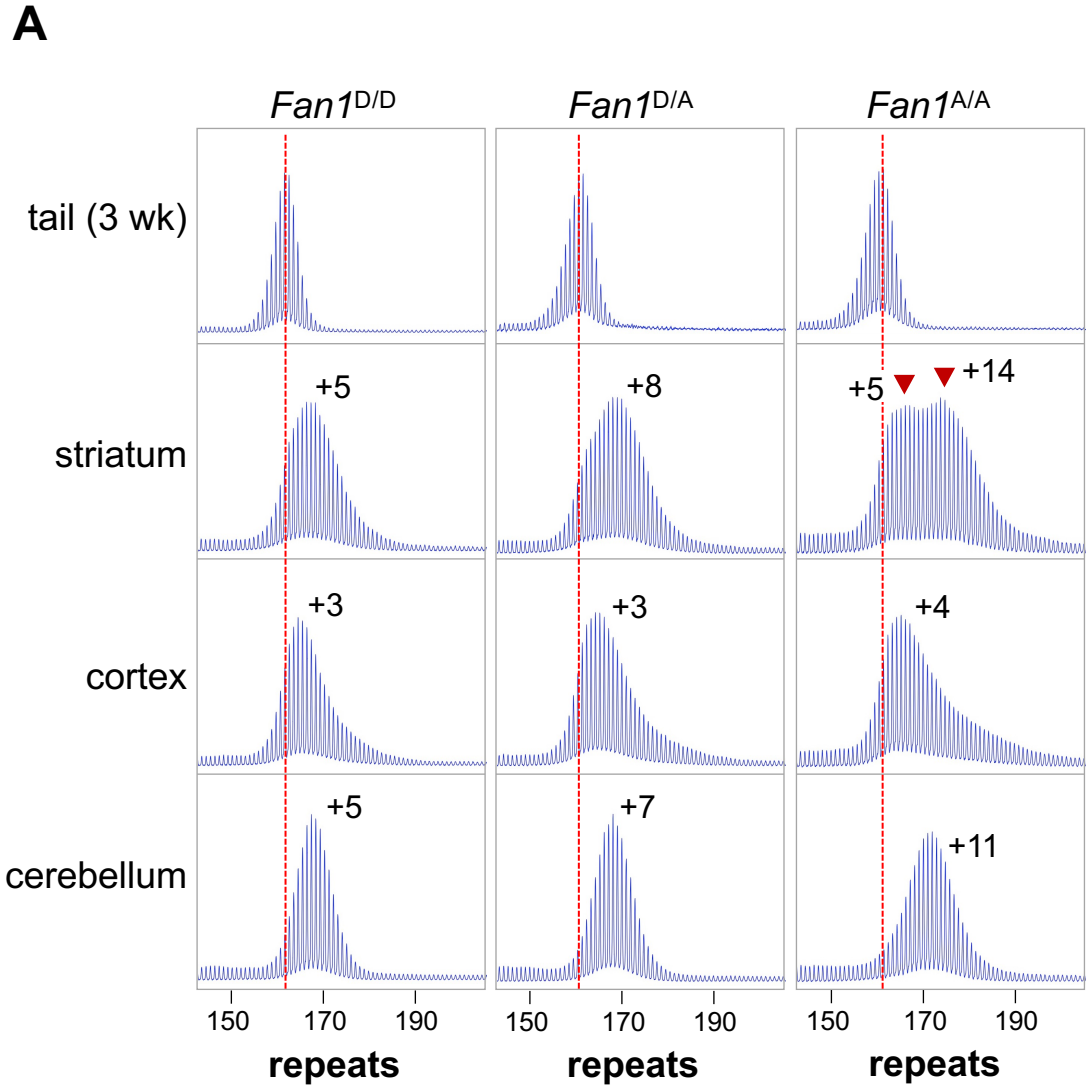
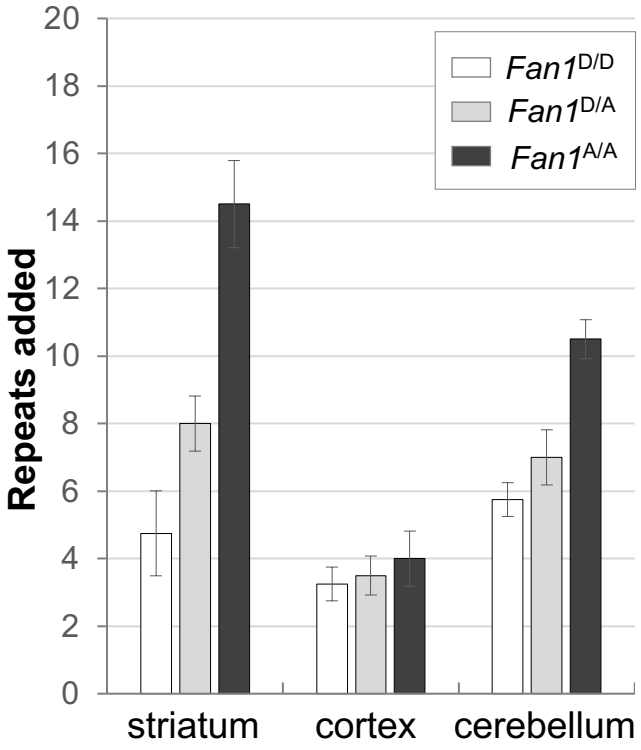
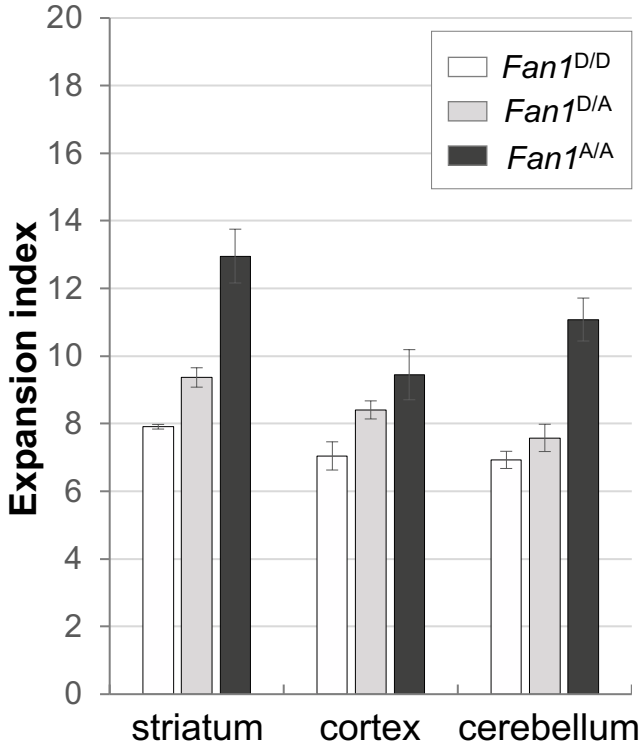

Relative Fan1 mRNA level
2.0
1.5
1.0
0.5
0.0
total brain
striatum
cortex
cerebellum

Fig.S7







Genotype	striatum	cortex	cerebellum
Fan1$^{D/D}$ vs *Fan1*$^{D/A}$	0.0160	0.7969	0.1024
Fan1$^{D/D}$ vs *Fan1*$^{A/A}$	<0.0001	0.3408	<0.0001
Fan1$^{D/A}$ vs *Fan1*$^{A/A}$	0.0008	0.6062	0.0016

Genotype	striatum	cortex	cerebellum
Fan1$^{D/D}$ vs *Fan1*$^{D/A}$	0.0029	0.0059	0.0886
Fan1$^{D/D}$ vs *Fan1*$^{A/A}$	0.0021	0.0066	0.0006
Fan1$^{D/A}$ vs *Fan1*$^{A/A}$	0.0030	0.124	0.0005

Fig.S8

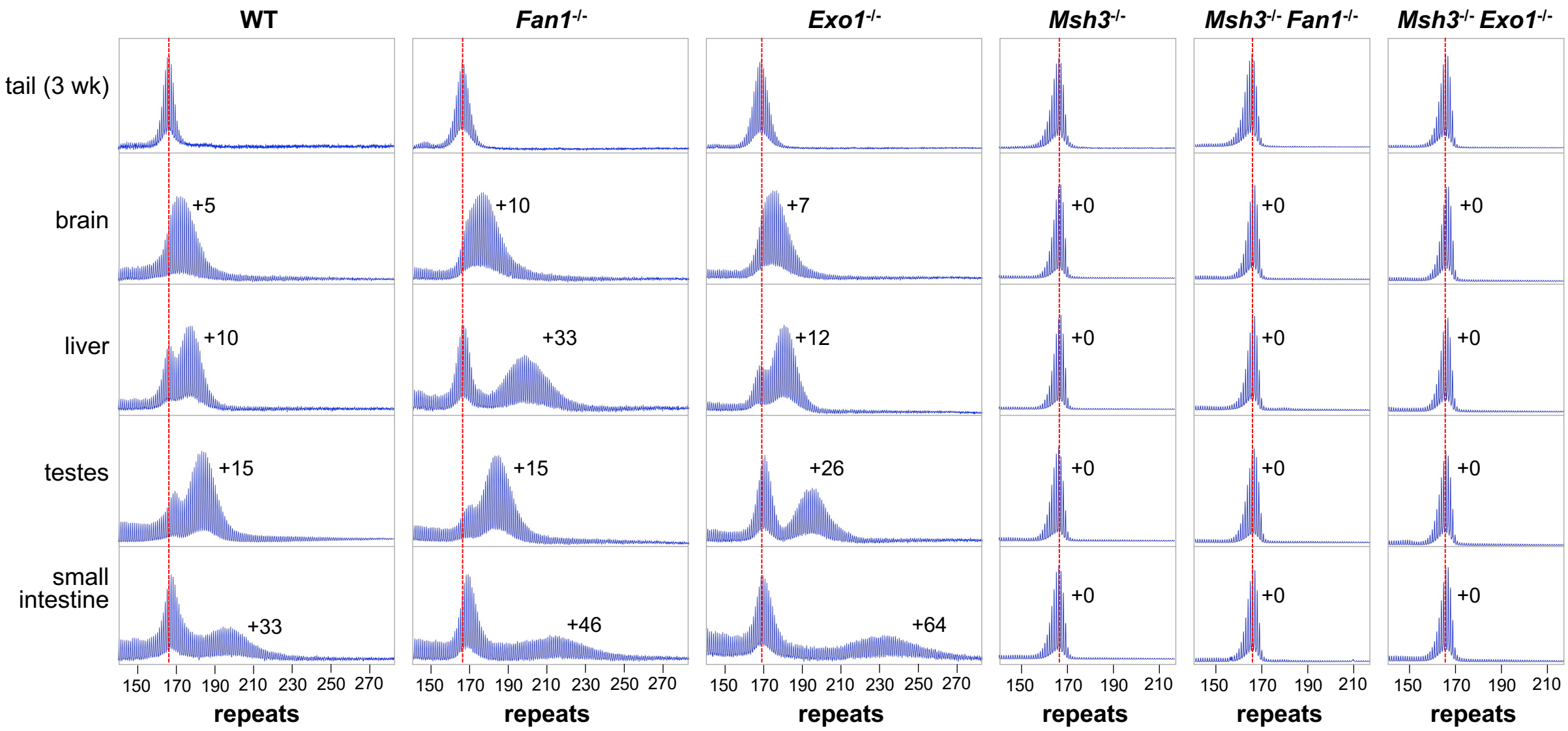

WT
Fan1-/-
Exo1-/-
Msh3-/-
Msh3-/- Fan1-/-
Msh3-/- Exo1-/-
tail (3 wk)
brain
liver
testes
small intestine
+5
+10
+7
+0
+0
+0
+10
+33
+12
+0
+0
+0
+15
+15
+26
+0
+0
+0
+33
+46
+64
+0
+0
+0
150 170 190 210 230 250 270
150 170 190 210
repeats

Fig. S9

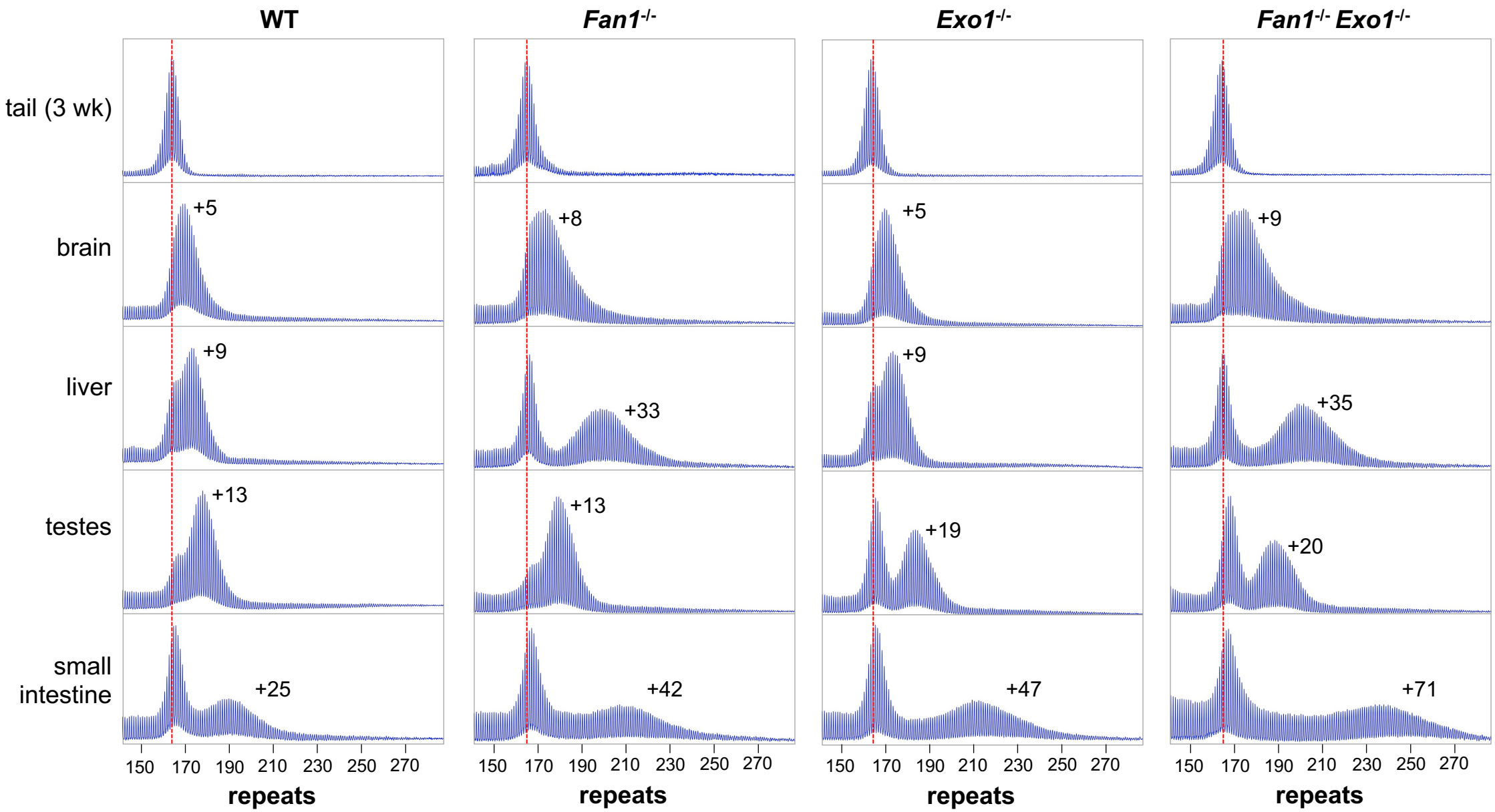

WT
Fan1-/-
Exo1-/-
Fan1-/- Exo1-/-
tail (3 wk)
brain
liver
testes
small intestine
+5
+8
+5
+9
+9
+33
+9
+35
+13
+13
+19
+20
+25
+42
+47
+71
150 170 190 210 230 250 270
repeats

Fig. S10

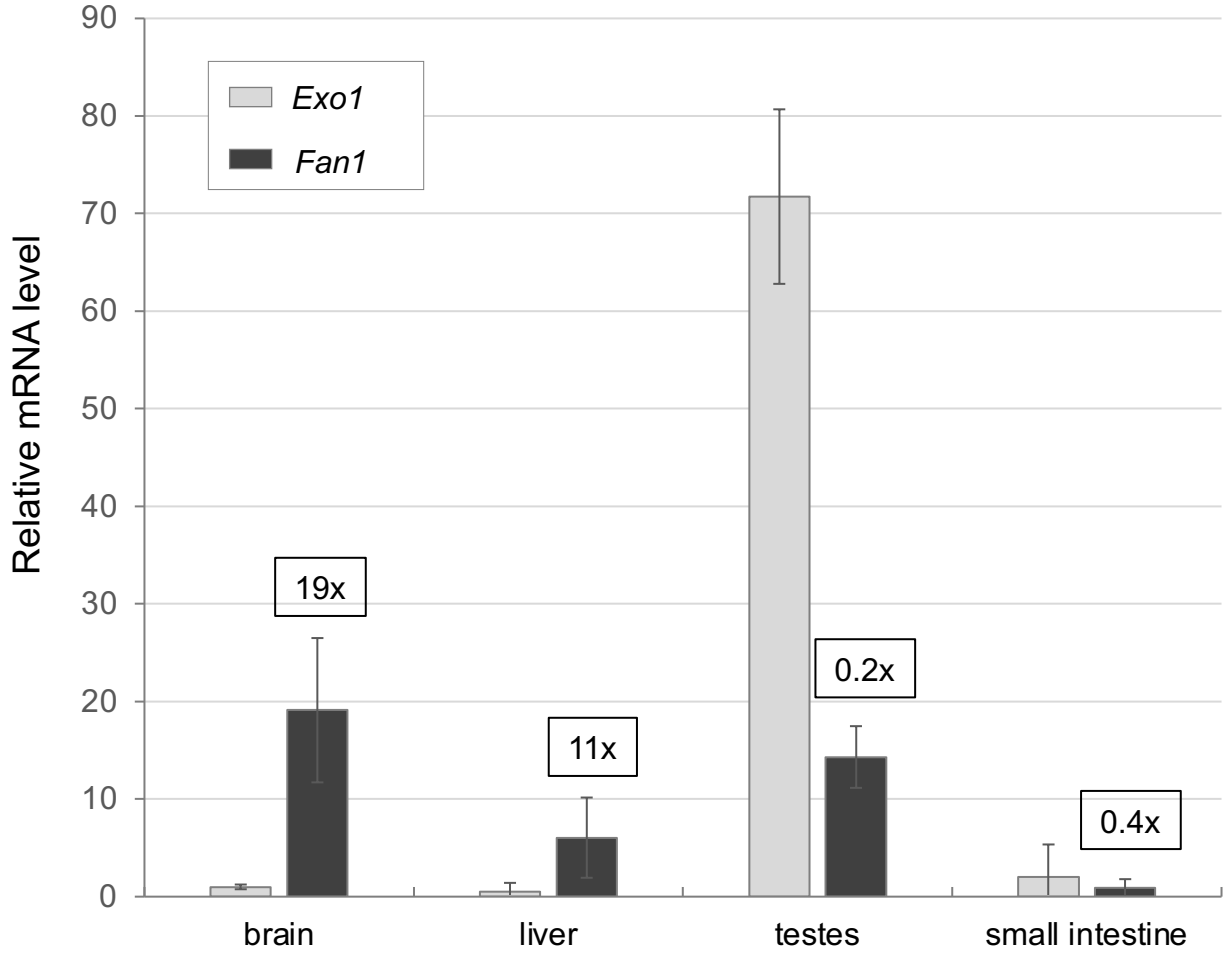

Relative mRNA level
90
80
70
60
50
40
30
20
10
0
Exo1
Fan1
19x
11x
0.2x
0.4x
brain
liver
testes
small intestine

Fig.S11