All three MutL complexes are required for repeat expansion in a human stem cell model of CAG-repeat expansion mediated glutaminase deficiency.

Bruce Hayward[1], Daman Kumari[1], André B.P. van Kuilenburg[2,3], and Karen Usdin[1,*]

[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] Amsterdam UMC location University of Amsterdam, Department of Clinical Chemistry, Laboratory Genetic Metabolic Diseases, Amsterdam, The Netherlands
[3] Amsterdam Gastroenterology Endocrinology Metabolism, Amsterdam, The Netherlands

*Corresponding author:

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, USA
email: karenu@nih.gov

Running title: The role of MutL proteins in GDPAG repeat expansion.

Keywords: Repeat expansion disease, microsatellite instability, mismatch repair (MMR), GDPAG or glutaminase deficiency disorder

Abstract

The Repeat Expansion Diseases (REDs) arise from expansion of a disease-specific short tandem repeat (STR). Different REDs differ with respect to the repeat involved, the cells that are most expansion prone and the extent of expansion and whether these diseases share a common expansion mechanism is unclear. To date, expansion has only been studied in a limited number of REDs. Here we report the first studies of the expansion mechanism in induced pluripotent stem cells derived from a patient with a form of the glutaminase deficiency disorder known as Global Developmental Delay, Progressive Ataxia, And Elevated Glutamine (GDPAG; OMIM# 618412) caused by the expansion of a CAG-STR in the 5' UTR of the glutaminase (*GLS*) gene. We show that alleles with as few as ~100 repeats show detectable expansions in culture despite relatively low levels of R-looped formed at this locus. Additionally, using a CRISPR-cas9 knockout approach we show that PMS2 and MLH3, the constituents of MutLα and MutLγ, the 2 mammalian MutL complexes known to be involved in mismatch repair (MMR), are essential for expansion. Furthermore, PMS1, a component of a less well understood MutL complex, MutLβ, is also important, if not essential, for repeat expansion in these cells. Our results provide insights into the factors important for expansion and lend weight to the idea that, despite some differences, many, if not all, REDs may expand in very similar ways.

Introduction

Global Developmental Delay, Progressive Ataxia, And Elevated Glutamine (GDPAG; OMIM# 618412) is a severe neurological disorder resulting from mutations in the *Glutaminase* (*GLS*) gene, a gene encoding the enzyme required for conversion of glutamine to glutamate, a critical excitatory neurotransmitter in the brain. Recently it has been demonstrated that some GDPAG cases are caused by expansion of a CAG-tandem repeat tract located in the 5' untranslated region of the *Glutaminase* (*GLS*) gene (1). More than 50 other human genetic disorders have been identified that also result from expansions of short tandem repeats (STRs) or microsatellites (2). These disorders are known collectively as the Repeat Expansion Disorders (REDs).

The REDs are caused by different STRs. They also differ in the cell types most prone to expansion and the extent of repeat expansion, with the number of repeats added varying widely from just a few repeats in the case of Huntington's disease (HD) to hundreds or even thousands of repeats in the case of other disorders like Myotonic Dystrophy Type 1 (DM1) (2). Whether the same mechanism is responsible for expansions in all REDs is still the subject of debate. While no study of the expansion mechanism in GDPAG has been reported to date, expansion has been studied in models of other REDs including HD, DM1, the Fragile X related disorders (FXDs) and Friedreich's ataxia (FRDA), disorders involving expansion of CAG, CTG, CGG and GAA-STRs respectively. Unlike the microsatellite instability (MSI) associated with cancer predisposition, mouse and cell models of the REDs suggest that some mismatch repair (MMR) proteins are actually required for expansion in these diseases (3-5), and recent genome wide association studies (GWAS) and

analysis of single nucleotide polymorphisms (SNPs) have shown that some of these same factors are modifiers of both somatic expansion risk and disease severity in patient cohorts (6-9).

The MMR factors known to be important for expansion in most model systems include the MutSβ mismatch recognition complex and MutLγ, one of the two MLH1-binding complexes known to be involved in mismatch processing in MMR (10). We had previously shown that in addition to MutLγ, a heterodimer of MLH1 and MLH3, that the second mismatch processing complex, MutLα, a heterodimer of MLH1 and PMS2, was also required for expansion in a mouse embryonic stem cell (mESC) model of the FXDs (11). Furthermore, MutLβ, a heterodimer of MLH1 and PMS1, a relatively abundant complex whose function is unclear, is also important, perhaps essential for expansion in these cells (11). The fact that all three MutL complexes are important for expansion was surprising since it suggested a hitherto unknown cooperation between these factors. Furthermore, an essential role for PMS2 was inconsistent with the demonstration that in a mouse model of DM1, loss of PMS2 resulted in a loss of only ~50% of expansions in the somatic tissues examined (12), while in a mouse model of FRDA, PMS2 actually suppressed expansions (13). This raised the question of whether the results we obtained in the mouse ESC model system were disease relevant not only for the FXDs, but for other REDs as well.

To begin to address this question, we used a gene knockout approach to examine the role of PMS1, PMS2 and MLH3 in GDPAG patient-derived iPSCs. This allows us to establish not only whether these factors promote or protect against expansion but whether they are essential for either process. This is important since an essential contribution indicate that these complexes

cooperate to generate the expansions rather than acting redundantly. We show here that all 3 MLH1-binding partners are also all important for expansions in the GDPAG patient-derived iPSCs, with PMS2 and MLH3 both being essential and with PMS1 also playing a vital, perhaps essential, role. This has implications not only for our understanding of the expansion process but for the possibility of a common therapeutic approach to reducing repeat expansion in multiple REDs as well.

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 and Sanger sequencing of plasmid constructs was carried out by Psomagen (Rockville, MD). All restriction enzymes were from New England Biolabs (Ipswich, MA). Plasmid pSpCas9(BB)-2A-Puro (PX459) V2.0 was a gift from Feng Zhang (Addgene plasmid # 62988; http://n2t.net/addgene:62988) and carries Cas9 from *Streptococcus pyogenes* (14). Plasmid pCK002_U6-Sa-sgRNA(mod)_EFS-SaCas9-2A-Puro_WPRE was a gift from Aviv Regev (Addgene plasmid # 85452; http://n2t.net/addgene:85452) and carries Cas9 from *Staphylococcus aureus* (15). pCE-mp53DD, carrying a dominant-negative p53 fragment that improves cell survival after CRISPRing, was a gift from Shinya Yamanaka (Addgene plasmid # 41856; http://n2t.net/addgene:41856) (16).

Plasmid construction

pXSaCas9-Puro2 was created by excising the gRNA scaffold, the SpCas9 promoter and the SpCas9 ORF of pX459v2 with *Bbs*I + *Bst*EII digestion and replacing them with an equivalent PCR fragment containing the SaCas9 gRNA scaffold, the SaCas9 promoter and the SaCas9 ORF from Addgene plasmid 85452 using a Gibson Assembly strategy. Sanger sequencing was performed to ensure the fidelity of the inserted PCR product. p53DD-alone was derived from pCE-mp53DD by *Hind*III + *Bam*HI digestion and recircularization to remove the EBNA1 and oriP sequences to prevent potential episomal propagation of the plasmid. Suitable gRNAs with minimal predicted off-targets for *MLH3* and *PMS1* were identified using the Synthego knockout guide algorithm (https://design.synthego.com/#/). Because of the complex nature of the *PMS2* locus (17), no acceptable SpCas9 gRNAs could be identified using either the Synthego algorithm or the Benchling (benchling.com) prediction tools. However, using the CRISPOR suite (http://crispor.tefor.net/, (18)) a suitable gRNA for SaCas9 that had minimal predicted off-targets was identified. Oligonucleotide pairs encoding these gRNAs were synthesized using MLH3_gRNA-F (CACCGTTGCTGTTGATGTAAGCAG) and MLH3_gRNA-R (AAACCTGCTTACATCAACAGCAAC); PMS1_gRNA-F (CACCGACATTTTATAGGAGAACTA) and PMS1_gRNA-R (AAACTAGTTCTCCTATAAAATGTC) and PMS2_gRNA-F (CACCGTCAAACCGTACTCTTCACACA) and PMS2_gRNA-R (GAACTGTGTGAAGAGTACGGTTTGAC) for *MLH3*, *PMS1* and *PMS2* respectively. The oligonucleotide pairs were annealed and cloned into either pX459v2 (*MLH3* and *PMS1*) or pXSaCas9-Puro2 (*PMS2*) which had been digested with *Bbs*I. All gRNA-encoding plasmids were verified by Sanger sequencing.

Cell culture

The derivation of iPSCs was carried out by the NHLBI iPSCs core facility using the Cytotune 2.0 Sendai virus kit (ThermoFisher Scientific, A16517) with a few modifications for high throughput reprogramming essentially as described earlier (19). Fibroblasts from an individual carrying a CAG expansion in the 5'-UTR of the GLS locus on both alleles (proband P2 in (1)) were used as the cell source. Pluripotency was assessed by staining for Tra-160, SSEA4 and Nanog by FACS analysis as described earlier (19). The iPSC cells were routinely maintained feeder-free in StemFlex media (ThermoFisher, A3349401) in wells coated with laminin-521 (StemCell Technologies, 200-0117). To passage the cells StemPro Accutase (ThermoFisher, A11105-01) was used to dissociate them and, after transfer to a fresh well the StemFlex was supplemented with RevitaCell (ThermoFisher, A2644501) for 24 hours. The iPSCs were made Sendai virus free by keeping in culture and testing for the presence of Sendai Virus (SeV) at different times by measuring SeV levels in total RNA using RT-qPCR with FAM labeled SeV (Thermofisher Scientific, assay ID Mr04269880_mr; 4331182) and VIC-labeled hGUSB Taqman probe/primers (Thermofisher Scientific, 4326320E) as the normalizing control.

DNA:RNA immunoprecipitation (DRIP) assay

DNA was isolated from cells using the salting out method (Miller, S.A. NAR 1988). DRIP was performed in duplicate for each sample using the anti RNA:DNA hybrid antibody (S9.6, Millipore-Sigma # MABE1095). For each DNA sample, three DRIPs were performed: normal mouse IgG (Millipore-Sigma # 12-371), S9.6 antibody without RNAse H treatment and RNAse H treatment followed by S9.6 antibody. A total of 25 µg DNA was either mock digested or digested with 30 units of RNAse H (New England Biolabs, Ipswich, MA) in 100 µl final volume at 37 °C for 6 hours.

The volume was made up to 400 µl with 300 µl of ChIP dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl, pH 8.1, 167 mM NaCl) and sonicated to fragments <500 bp using Bioruptor® (6 minutes at medium setting, 30 seconds ON/30 seconds OFF). To 350 µl of the sonicated DNA, 650 µl of the ChIP dilution buffer and 10 µl of protease inhibitor cocktail (Sigma-Aldrich, St. Louis, MO) was added and mixed. An aliquot (1%) was saved as input sample. The sonicated DNA was then precleared with 50 µl of Protein A agarose beads/Salmon sperm DNA slurry (EMD Millipore, Catalog # 16–157) for one hour on a rotator in cold. The precleared supernatant was incubated with 5 µg of mouse IgG or S9.6 antibody overnight on a rotator in a cold room. The sample was then incubated with 60 µl of the Protein A agarose beads/Salmon sperm DNA slurry to collect the immune complexes. The material was washed once for 5 minutes each with low salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.1, 150 mM NaCl), high salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.1, 500 mM NaCl), LiCl wash buffer (0.25 M LiCl, 1% IGEPAL-CA630, 1% deoxycholic acid (sodium salt), 1 mM EDTA, 10 mM Tris, pH 8.0), and twice with TE pH 8.0 (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). The immunoprecipitated material was eluted from the beads using 2 x 250 µl of elution buffer (1% SDS, 100 mM NaHCO3). To the input and eluted DNA samples, 250 nM NaCl and 200 nM NaCl, respectively, was added and the samples incubated at 65°C for 4.5 hours to allow reverse crosslinking. The input and DRIP samples were then treated with phenol/chloroform and precipitated overnight at −20 °C with 0.3 M sodium acetate and ethanol in the presence of pellet paint. After washing with 70% ethanol, the samples were resuspended in 50 µl 0.1X TE pH 8.0. Real-time PCRs on the immunoprecipitated DNAs were carried out in triplicate in 20 µl final volume using the PowerUP™ SYBR™ Green PCR master mix (ThermoFisher

Scientific) and 2 µl of DNA using QuantStudio™ 3 real-time PCR system (ThermoFisher Scientific). For the amplification of *GLS* promoter region upstream of the repeat, 150 nM each of primer GLS-F2 (5' GATTTGAGCCAATCGCAGC 3') and GLS-R1 (5' GGCTAGAGACCTTCAGCGCT 3') were used and for the amplification of GLS exon1 region downstream of the repeat, 300 nM each of primer GLS Ex1-F (5'-cccaagtagctgccctttcc-3') and GLS Ex1-R (5'-cgctcaacaggggaggatg-3') were used. For the amplification of positive control *FMR1* exon1 region, 300 nM each of *FMR1* Exon1-F (5'-CGCTAGCAGGGCTGAAGAGAA-3') and *FMR1* Exon1-R (5'-GTACCTTGTAGAAAGCGCCATTGGAG-3') were used. For the amplification of negative control region *ZNF554*, 300 nM each of *ZNF554*-F (5'-CGGGGAAAAGCCCTATAAAT-3') and *ZNF554*-R (5'-TCCACATTCACTGCATTCGT-3') were used.

Generation of CRISPR knockouts

To generate the desired CRISPR knockouts ~100K iPSCs were plated into a 12-well 24 hours before transfection. Immediately prior to transfection the media was replaced with 0.5 ml of OptiMem (ThermoFisher, 31985062) supplemented with RevitaCell. Transfections were carried out using 800 ng of the appropriate Cas9 plasmid and 200 ng of p53DD-alone complexed with 4µl of Lipofectamine Stem (ThermoFisher, STEM00008) as per the manufacturer's instructions. Four hours after the transfection, 0.5 ml of StemFlex supplemented with RevitaCell was added to the cells. Approximately 24 hours after the transfection, the media was replaced with 1 ml StemFlex supplemented with RevitaCell containing 1 µg/ml Puromycin and selection applied for 24-48 hours. The culture was allowed to recover for 4 days and then ~4K cells were plated into a 60mm dish to allow single cell derived colonies to appear. 24 colonies were picked, genomic DNA

prepared using the KAPA Express reagent (a component of the KAPA Mouse Genotyping Kit from Roche, 07961804001) and analyzed for both CRISPR lesions and repeat size as described below.

The lesions present in the *MLH3* and *PMS1* lines were identified by direct sequencing of the PCR product containing the gRNA target and subsequent analysis by the ICE tool from Synthego (ice.synthego.com). Sequencing of the cDNA was used to confirm the loss of the entire target exon in one *MLH3* clone that did not generate a genomic PCR product. The lesions present in the individual *PMS2* cell lines were identified by sequencing of individual plasmid clones derived from the PCR product containing the gRNA target. A known SNP (rs1805319) allowed the 2 alleles to be unambiguously identified. The mutations generated are shown in Table 1.

Analysis of repeat size

The GLS repeats were amplified using the primer pair GLS-F and GLS-R, as previously described (1) except that GLS-R was 5'-FAM labelled. The PCR reaction (15 µl) contained 50 mM Tris-HCl pH 9.0, 1.75 mM $MgCl_2$, 22 mM $(NH_4)_2SO_4$, 0.5 µM each primer, 1.5 M betaine, 2% DMSO and 0.2 mM each dATP, dCTP, dGTP and dTTP. Three tenths of a microliter of Robust HotStart KAPA2G polymerase (Roche, 07961316001) was added per 100 µl of PCR mix. Cycling parameters of 98°C for 3 minutes, 35 cycles of (98°C for 30 seconds, 65°C for 30 seconds, 72°C for 210 seconds), and 72°C for 10 minutes were used. The samples were then resolved by capillary electrophoresis. The fragment analysis data files from the capillary electrophoresis were analyzed using a custom R script as previously described (20).

Results

Repeat expansions are seen in GDPAG patient derived iPSCs.

We isolated iPSCs from fibroblasts from an individual with expansions in both *GLS* alleles (Patient P2 in (1)). These fibroblasts contained a heterogeneous mixture of alleles of different sizes and, as a result, iPSC lines with a range of different repeat sizes were obtained, most having ~600->1000 repeats in each allele. As can be seen from PCR analysis of the repeat over time in culture, expansions similar to those reported in iPSCs from other REDs (6,21-23) could be detected by agarose gel electrophoresis (Fig. S1). Specifically, the alleles in the population appear to expand in concert, a phenomenon consistent with the high frequency addition of a small number of repeats (24). While large alleles are not well resolved by agarose gel electrophoresis, thus precluding an accurate determination of repeat number, we estimate that larger alleles in the GLS C and GLD D lines gained approximately 60 repeats over a period of 50-59 days (Fig. S1), corresponding to the gain of ~1 repeat/day or ~1 expansion event/day in the majority of cells in the population.

Most of these alleles are too large to be analyzed by capillary electrophoresis (CE), a simple cost-effective, high-resolution technique that allows alleles differing by one repeat to be reliably distinguished (20). However, one iPSC line (GLS B) had sustained a contraction resulting in an allele with ~104 repeats, putting it in the range suitable for CE analysis. We then derived additional cell lines from this one by plating the iPSC line for single colonies. This resulted in individual cell lines with repeat numbers that varied by 1-4 repeats. Six of these lines were grown for 49 days. As can be seen from Fig. 1 and Fig. 2, the allele in two of the 6 lines gained 7 repeats

over 49 days in culture as evidenced by a clear shift of the modal allele by 7 repeats. In the case of the remaining cell lines the two largest allele populations were very similar in abundance making it impossible to identify a single modal allele. Since these allele sizes corresponded to a gain of 6 and 7 repeats respectively, it suggests that these cell lines gained on average ~6.5 repeats over the 49 days in culture. Thus, these lines are all expanding at a rate of ~0.14 repeats/day or ~ 1 repeat/week. In contrast, we have previously shown that two FXD iPSC lines with similar number of CGG-repeats in the *FMR1* locus did not show evidence of expansion over a similar span of time (25).

R-loop formation at the GLS locus

R-loops, three-stranded nucleic acid structures containing an RNA:DNA hybrid, are commonly seen at disease-associated STRs, where they have been suggested to play a role in the expansion process (10,26-29). To explore this idea in the context of *GLS* expansion, we used the S9.6 antibody (30) to assess the levels of R-loop formed in iPSCs from an unaffected individual and the GDPAG iPSCs using *ZNF554* and *FMR1* as negative and positive controls respectively (31-34). As can be seen in Fig. 3, both the iPSC lines produced a similar DRIP signal at the *GLS* promoter and exon 1 of the gene, despite the fact that GLS mRNA levels in the patient iPSCs are more than 4-fold lower (data not shown). The DRIP signals seen at both *GLS* regions tested were significantly lower than the DRIP signal at *FMR1* in both the unaffected and patient lines and not significantly different from the DRIP signals seen for the negative control in these lines. This suggests that R-loop formation may not be a major factor in expansion at the *GLS* locus.

Null mutations in PMS1, PMS2 and MLH3 eliminate almost all expansions in the GLSD iPSCs.

Using CRISPR-Cas9 we made null mutations in *PMS1*, *PMS2* and *MLH3* as described in the Materials and Methods. We isolated 3 independent sequence-verified null lines for each gene. The mutations generated in each cell line are listed in Table 1. In the case of the *PMS1* null lines, 2 lines had repeat numbers within the range of the unedited lines, with one cell line having 1 repeat fewer than the smallest unedited line. All of the *PMS2* and *MLH3* null lines had repeat numbers within the range of the unedited lines. We then grew each of these lines in culture for 49 days (Fig 4 – 6). As can be seen from Fig. 4, each of the *PMS1* null lines showed a subtle shift in their PCR profile that resulted in the modal allele being one repeat larger than the modal allele in the original culture. This represents at least an 86% reduction in the extent of expansion compared to unedited lines. The number of repeats in the *PMS2* and *MLH3* null lines did not change over the same period (Fig 5 - 6), consistent with the idea that MutLα and MutLγ are both required for expansion in these cells.

Discussion

Here we demonstrate that GDPAG patient-derived iPSCs constitute a tractable human model system to study "expansion in a dish", with alleles as small as ~100 repeats showing clear evidence of expansion over an experimental reasonable timeframe. This allows the accurate measurement of changes in repeat number at single repeat resolution and the ready evaluation of factors that impact expansion. However, while we saw expansion in these cells (Figs 1-2), we saw little evidence of R-loop formation in cells heterozygous for this allele and an allele that is 10 times larger (Fig. 3). This casts doubt on a role for R-loops in promoting expansion.

We also used these cells to examine the effect of CRISPR-generated null mutations in *PMS1*, *PMS2* or *MLH3*, genes we had previously shown to be required for expansions in mESCs from a mouse model of the FXDs (11). The effect of the loss of each of these proteins was broadly similar to what we found in the mESC model (11). Specifically, in the case of *PMS2* and *MLH3*, no repeats were gained. This is consistent with the idea that both MLH1-binding partners are essential for expansion in these cells. However, unlike what was seen with the FXD mESCs where 1-2 repeats were lost over a similar period, no repeats were lost in this case. We had previously attributed the loss of repeats in the FXD mESCs to the effect of the MLH3 or PMS2-deficiency on normal MMR. The failure to see a similar loss in the GDPAG-patient iPSCs may reflect the smaller number of repeats in these lines. This might make the repeat tract less prone to any microsatellite instability resulting from an MMR-deficiency. As we saw with the FXD mESCs, loss of PMS1 resulted in the growth of the repeat tract being limited to a single repeat, in this case corresponding to an 86% decrease in expansions over a 7-week period. Whether the residual repeat gained reflects a small fraction of PMS1-independent expansion events, or some effect of the absence of PMS1 on normal MMR is unclear.

Our results clearly demonstrate that the contribution of all three MutL proteins to expansion is not unique to the mouse model of the FXDs. Rather it may be a general feature of expansion in the REDs. This raises the question of whether PMS2's paradoxical effects, namely its ability to protect against expansions in some model systems and its requirement for expansion in others, may not reflect different mechanisms for generating expansions in different diseases but rather

differences related to cell type and/or the sequence or location of the repeat in the genome. For example, as illustrated in Fig. S2, it is possible that PMS2's promutagenic and antimutagenic effects on expansion are related to the relative levels of MLH3 and PMS2 in the cell type being evaluated. Our finding that PMS1 is also involved in generating expansions in the case of repeat-induced glutaminase deficiency lends weight to the idea that PMS1 may be a common target for therapies aimed at reducing somatic expansion in multiple REDs.

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).

Conflict of interest statement

The authors have no conflicts to declare.

Acknowledgements

We would like to thank Carson Miller for help with changing media over the weekends and the members of the Usdin lab for insightful discussions.

References

1. van Kuilenburg, A.B.P., Tarailo-Graovac, M., Richmond, P.A., Drogemoller, B.I., Pouladi, M.A., Leen, R., Brand-Arzamendi, K., Dobritzsch, D., Dolzhenko, E., Eberle, M.A. *et al.* (2019) Glutaminase Deficiency Caused by Short Tandem Repeat Expansion in GLS. *N Engl J Med*, **380**, 1433-1441.
2. Paulson, H. (2018) Repeat expansion diseases. *Handb Clin Neurol*, **147**, 105-123.
3. Zhao, X. and Usdin, K. (2021) (Dys)function Follows Form: Nucleic Acid Structure, Repeat Expansion, and Disease Pathology in FMR1 Disorders. *Int J Mol Sci*, **22**.
4. Zhao, X., Gazy, I., Hayward, B., Pintado, E., Hwang, Y.H., Tassone, F. and Usdin, K. (2019) Repeat Instability in the Fragile X-Related Disorders: Lessons from a Mouse Model. *Brain Sci*, **9**.
5. Wheeler, V.C. and Dion, V. (2021) Modifiers of CAG/CTG Repeat Instability: Insights from Mammalian Models. *J Huntingtons Dis*, **10**, 123-148.
6. McAllister, B., Donaldson, J., Binda, C.S., Powell, S., Chughtai, U., Edwards, G., Stone, J., Lobanov, S., Elliston, L., Schuhmacher, L.N. *et al.* (2022) Exome sequencing of individuals with Huntington's disease implicates FAN1 nuclease activity in slowing CAG expansion and disease onset. *Nat Neurosci*, **25**, 446-457.
7. Kim, K.H., Hong, E.P., Shin, J.W., Chao, M.J., Loupe, J., Gillis, T., Mysore, J.S., Holmans, P., Jones, L., Orth, M. *et al.* (2020) Genetic and Functional Analyses Point to FAN1 as the Source of Multiple Huntington Disease Modifier Effects. *Am J Hum Genet*, **107**, 96-110.
8. Hwang, Y.H., Hayward, B.E., Zafarullah, M., Kumar, J., Durbin Johnson, B., Holmans, P., Usdin, K. and Tassone, F. (2022) Both cis and trans-acting genetic factors drive somatic instability in female carriers of the FMR1 premutation. *Scientific Reports*, **12**, 10419.
9. Flower, M., Lomeikaite, V., Ciosi, M., Cumming, S., Morales, F., Lo, K., Hensman Moss, D., Jones, L., Holmans, P., Investigators, T.-H. *et al.* (2019) MSH3 modifies somatic instability and disease severity in Huntington's and myotonic dystrophy type 1. *Brain*, **142**, 1876-1886.
10. Lin, Y., Leng, M., Wan, M. and Wilson, J.H. (2010) Convergent transcription through a long CAG tract destabilizes repeats and induces apoptosis. *Mol Cell Biol*, **30**, 4435-4451.
11. Miller, C.J., Kim, G.Y., Zhao, X. and Usdin, K. (2020) All three mammalian MutL complexes are required for repeat expansion in a mouse cell model of the Fragile X-related disorders. *PLoS Genet*, **16**, e1008902.
12. Gomes-Pereira, M., Fortune, M.T., Ingram, L., McAbney, J.P. and Monckton, D.G. (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.
13. Bourn, R.L., De Biase, I., Pinto, R.M., Sandi, C., Al-Mahdawi, S., Pook, M.A. and Bidichandani, S.I. (2012) Pms2 suppresses large expansions of the (GAA.TTC)n sequence in neuronal tissues. *PLoS One*, **7**, e47085.
14. Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A. and Zhang, F. (2013) Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*, **8**, 2281-2308.

15. Singer, M., Wang, C., Cong, L., Marjanovic, N.D., Kowalczyk, M.S., Zhang, H., Nyman, J., Sakuishi, K., Kurtulus, S., Gennert, D. *et al.* (2016) A Distinct Gene Module for Dysfunction Uncoupled from Activation in Tumor-Infiltrating T Cells. *Cell*, **166**, 1500-1511 e1509.
16. Okita, K., Yamakawa, T., Matsumura, Y., Sato, Y., Amano, N., Watanabe, A., Goshima, N. and Yamanaka, S. (2013) An efficient nonviral method to generate integration-free human-induced pluripotent stem cells from cord blood and peripheral blood cells. *Stem Cells*, **31**, 458-466.
17. Clendenning, M., Hampel, H., LaJeunesse, J., Lindblom, A., Lockman, J., Nilbert, M., Senter, L., Sotamaa, K. and de la Chapelle, A. (2006) Long-range PCR facilitates the identification of PMS2-specific mutations. *Hum Mutat*, **27**, 490-495.
18. Concordet, J.P. and Haeussler, M. (2018) CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res*, **46**, W242-W245.
19. Beers, J., Linask, K.L., Chen, J.A., Siniscalchi, L.I., Lin, Y., Zheng, W., Rao, M. and Chen, G. (2015) A cost-effective and efficient reprogramming platform for large-scale production of integration-free human induced pluripotent stem cells in chemically defined culture. *Sci Rep*, **5**, 11319.
20. Hayward, B.E., Zhou, Y., Kumari, D. and Usdin, K. (2016) A Set of Assays for the Comprehensive Analysis of FMR1 Alleles in the Fragile X-Related Disorders. *J Mol Diagn*, **18**, 762-774.
21. Goold, R., Hamilton, J., Menneteau, T., Flower, M., Bunting, E.L., Aldous, S.G., Porro, A., Vicente, J.R., Allen, N.D., Wilkinson, H. *et al.* (2021) FAN1 controls mismatch repair complex assembly via MLH1 retention to stabilize CAG repeat expansion in Huntington's disease. *Cell Rep*, **36**, 109649.
22. Du, J., Campau, E., Soragni, E., Jespersen, C. and Gottesfeld, J.M. (2013) Length-dependent CTG.CAG triplet-repeat expansion in myotonic dystrophy patient-derived induced pluripotent stem cells. *Hum Mol Genet*, **22**, 5276-5287.
23. Ku, S., Soragni, E., Campau, E., Thomas, E.A., Altun, G., Laurent, L.C., Loring, J.F., Napierala, M. and Gottesfeld, J.M. (2010) Friedreich's ataxia induced pluripotent stem cells model intergenerational GAA·TTC triplet repeat instability. *Cell Stem Cell*, **7**, 631-637.
24. Møllersen, L., Rowe, A.D., Larsen, E., Rognes, T. and Klungland, A. (2010) Continuous and periodic expansion of CAG repeats in Huntington's disease R6/1 mice. *PLoS Genet*, **6**, e1001242.
25. Zhou, Y., Kumari, D., Sciascia, N. and Usdin, K. (2016) CGG-repeat dynamics and FMR1 gene silencing in fragile X syndrome stem cells and stem cell-derived neurons. *Mol Autism*, **7**, 42.
26. Su, X.A. and Freudenreich, C.H. (2017) Cytosine deamination and base excision repair cause R-loop-induced CAG repeat fragility and instability in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A*, **114**, E8392-E8401.
27. Neil, A.J., Liang, M.U., Khristich, A.N., Shah, K.A. and Mirkin, S.M. (2018) RNA-DNA hybrids promote the expansion of Friedreich's ataxia (GAA)n repeats via break-induced replication. *Nucleic Acids Res*, **46**, 3487-3497.
28. Khristich, A.N. and Mirkin, S.M. (2020) On the wrong DNA track: Molecular mechanisms of repeat-mediated genome instability. *J Biol Chem*, **295**, 4134-4170.

29. McIvor, E.I., Polak, U. and Napierala, M. (2010) New insights into repeat instability: role of RNA*DNA hybrids. *RNA Biol*, **7**, 551-558.
30. Boguslawski, S.J., Smith, D.E., Michalak, M.A., Mickelson, K.E., Yehle, C.O., Patterson, W.L. and Carrico, R.J. (1986) Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. *J Immunol Methods*, **89**, 123-130.
31. Li, L., Matsui, M. and Corey, D.R. (2016) Activating frataxin expression by repeat-targeted nucleic acids. *Nat Commun*, **7**, 10606.
32. Loomis, E.W., Sanz, L.A., Chedin, F. and Hagerman, P.J. (2014) Transcription-associated R-loop formation across the human FMR1 CGG-repeat region. *PLoS Genet*, **10**, e1004294.
33. Kumari, D. and Usdin, K. (2016) Sustained expression of FMR1 mRNA from reactivated fragile X syndrome alleles after treatment with small molecules that prevent trimethylation of H3K27. *Hum Mol Genet*, **25**, 3689-3698.
34. Abu Diab, M., Mor-Shaked, H., Cohen, E., Cohen-Hadad, Y., Ram, O., Epsztejn-Litman, S. and Eiges, R. (2018) The G-rich Repeats in FMR1 and C9orf72 Loci Are Hotspots for Local Unpairing of DNA. *Genetics*, **210**, 1239-1252.

Tables

Table 1. Description of different CRISPR-generated mutant lines

Line	Allele 1	Effect	Allele 2	Effect	Verification
PMS1 #1	-18 bp	splice site loss	ND		Synthego ICE
PMS1 #2	-11 bp	frame shift	-21 bp	splice site loss	Synthego ICE
PMS1 #3	-1 bp	frame shift	-10 bp	frame shift	Synthego ICE
PMS2 #1	-4 bp	frame shift	-4 bp	frame shift	sequencing
PMS2 #2	-1 bp	frame shift	+198 bp insert, -7 bp	in frame stop	sequencing
PMS2 #3	-4 bp	frame shift	-4 bp	frame shift	sequencing
MLH3 #1	-4 bp	frame shift	ND		Synthego ICE
MLH3 #2	-14 bp	frame shift	ND		Synthego ICE
MLH3 #3	ND	exon skip	ND		cDNA sequencing

Figure Legends

Fig. 1. Repeat PCR profiles of 6 clones derived from unedited patient iPSCs. Six clones with similar repeat numbers were grown for 49 days and samples removed for analysis at the indicated time points. In each case, the red dotted line indicates the repeat number present in the original culture. The numbers alongside each plot indicate the change in repeat number at that time-point.

Fig. 2. Graphical representation of the average change in repeat number over time in 6 unedited patient iPSC lines.

Fig. 3. DNA:RNA immunoprecipitation (DRIP) signal at the *GLS* promoter and exon 1 region in unaffected control and patient iPSCs. Data shown is an average of two independent experiments

and error bars represent standard deviation. The *FMR1* exon 1 was used as a positive control and the *ZNF554* gene was used as a negative control. Two-tailed, unpaired student t test was used to calculate p-values. The p-values shown indicate the comparison with the negative control. n.s., not significant.

Fig. 4. Repeat PCR profiles of 3 different *PMS1*$^{-/-}$ patient iPSC lines. Three lines with similar repeat numbers were grown for 49 days and samples removed for analysis at the indicated time points. In each case, the red dotted line indicates the repeat number present in the original culture. The numbers alongside the last plot indicate the change in repeat number at that time-point.

Fig. 5. Repeat PCR profiles of 3 different *PMS2*$^{-/-}$ patient iPSC lines. Three lines with similar repeat numbers were grown for 49 days and samples removed for analysis at the indicated time points. In each case, the red dotted line indicates the repeat number present in the original culture. The numbers alongside the last plot indicate the change in repeat number at that time-point.

Fig. 6. Repeat PCR profiles of 3 different *MLH3*$^{-/-}$ patient iPSC lines. Three lines with similar repeat numbers were grown for 49 days and samples removed for analysis at the indicated time points. In each case, the red dotted line indicates the repeat number present in the original culture. The numbers alongside the last plot indicate the change in repeat number at that time-point.

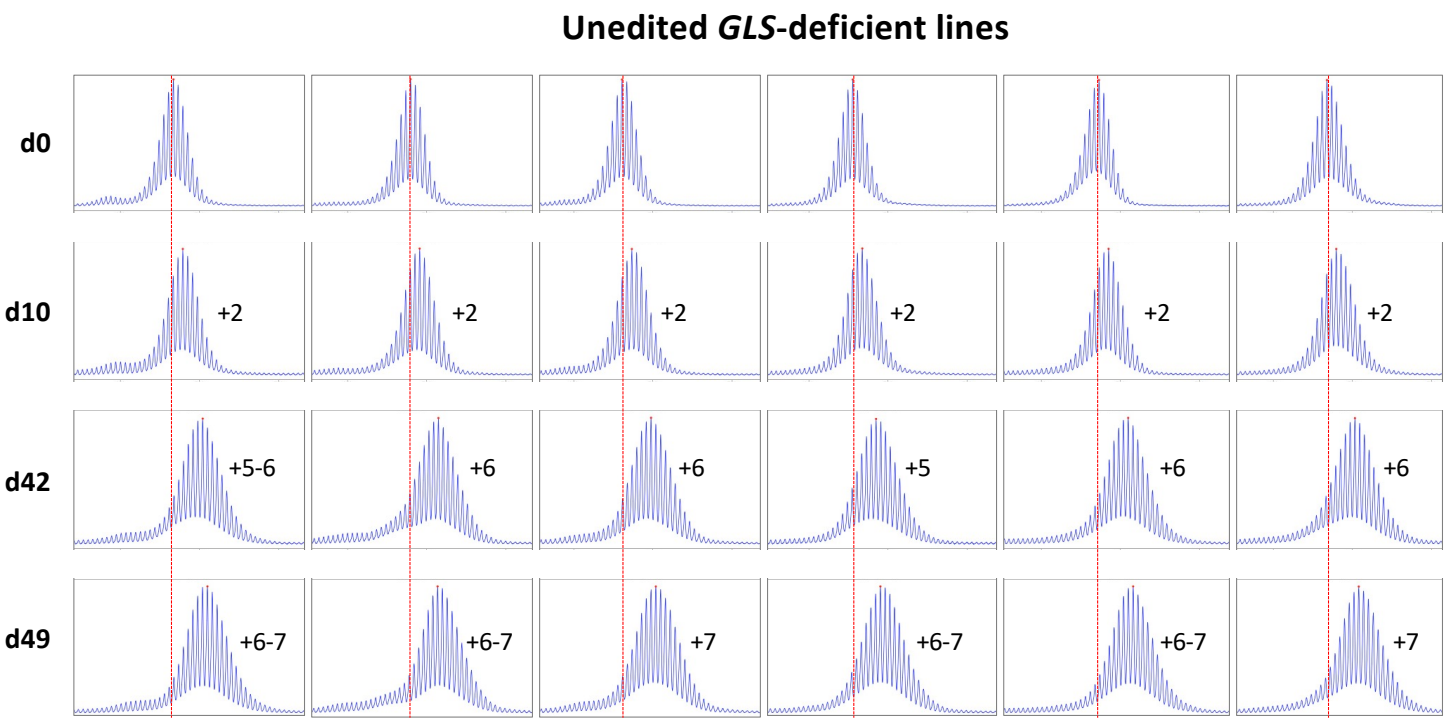
Unedited *GLS*-deficient lines
d0
d10
d42
d49
+2
+2
+2
+2
+2
+2
+5-6
+6
+6
+5
+6
+6
+6-7
+6-7
+7
+6-7
+6-7
+7

Fig. 1

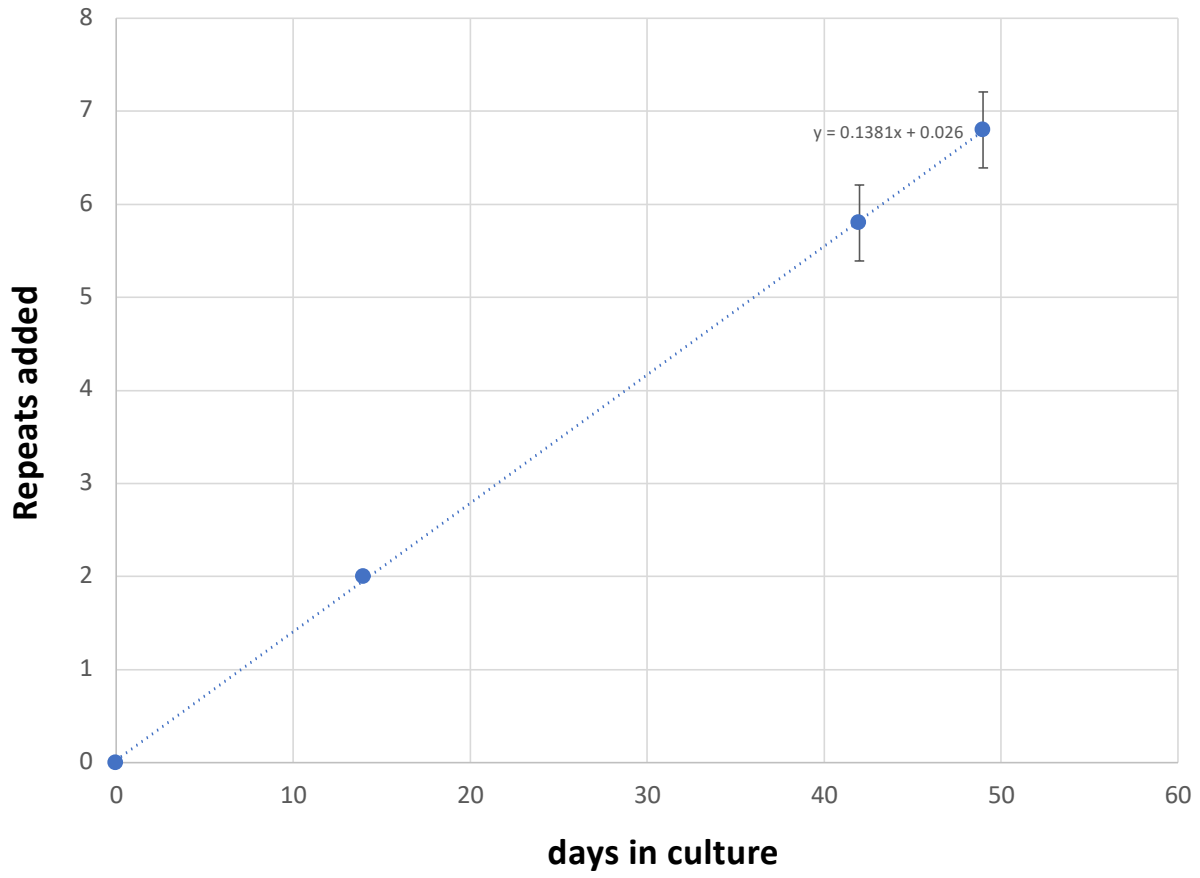
y = 0.1381x + 0.026
Repeats added
days in culture
0
1
2
3
4
5
6
7
8
0
10
20
30
40
50
60

Fig. 2

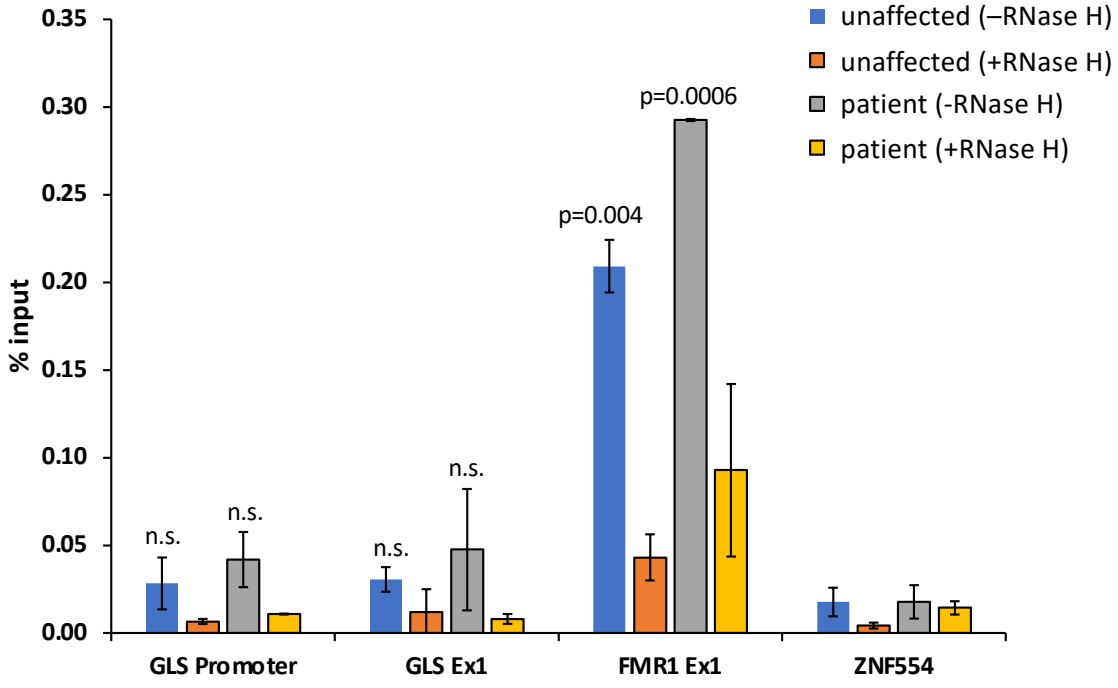
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
% input
unaffected (–RNase H)
unaffected (+RNase H)
patient (-RNase H)
patient (+RNase H)
p=0.0006
p=0.004
n.s.
n.s.
n.s.
n.s.
GLS Promoter
GLS Ex1
FMR1 Ex1
ZNF554

Fig. 3

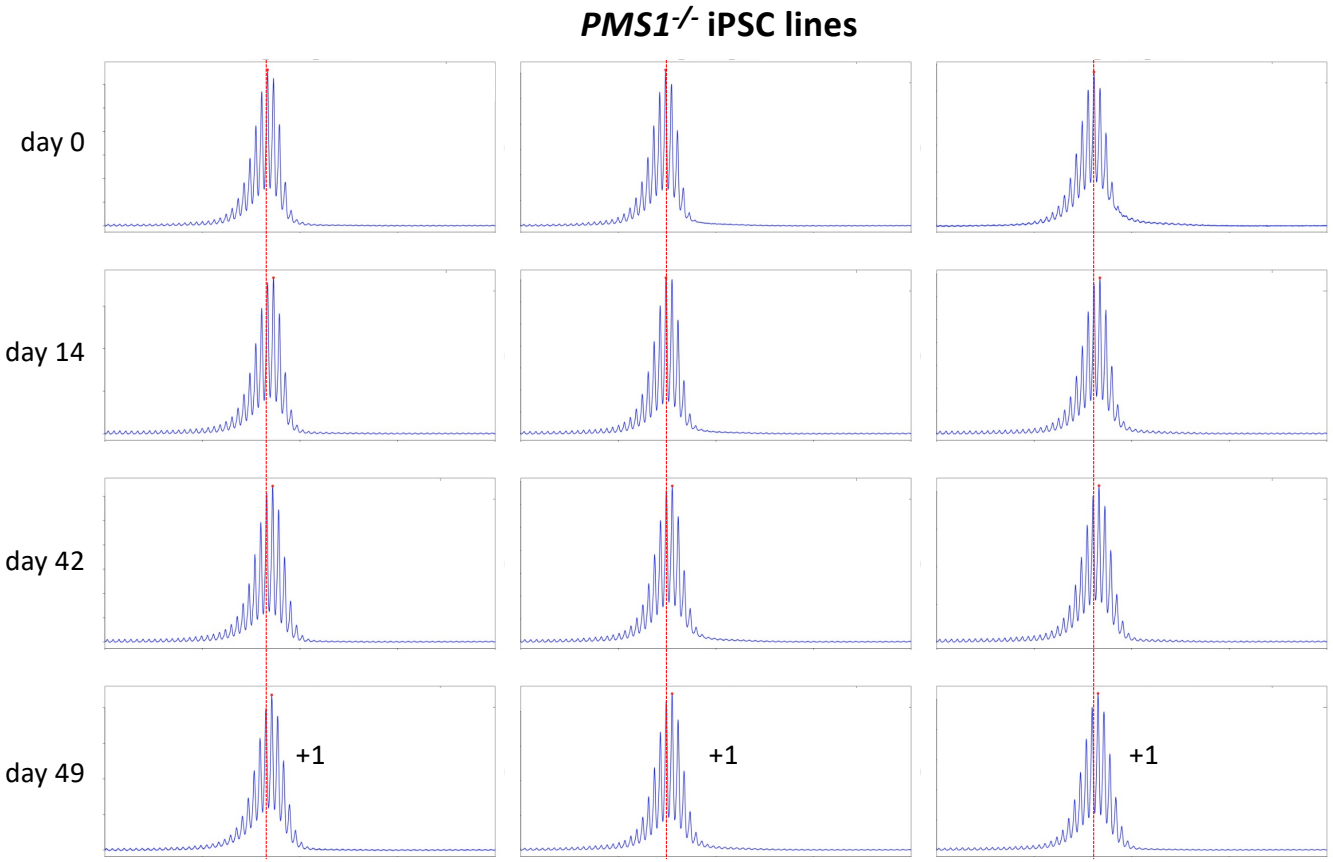

Fig. 4

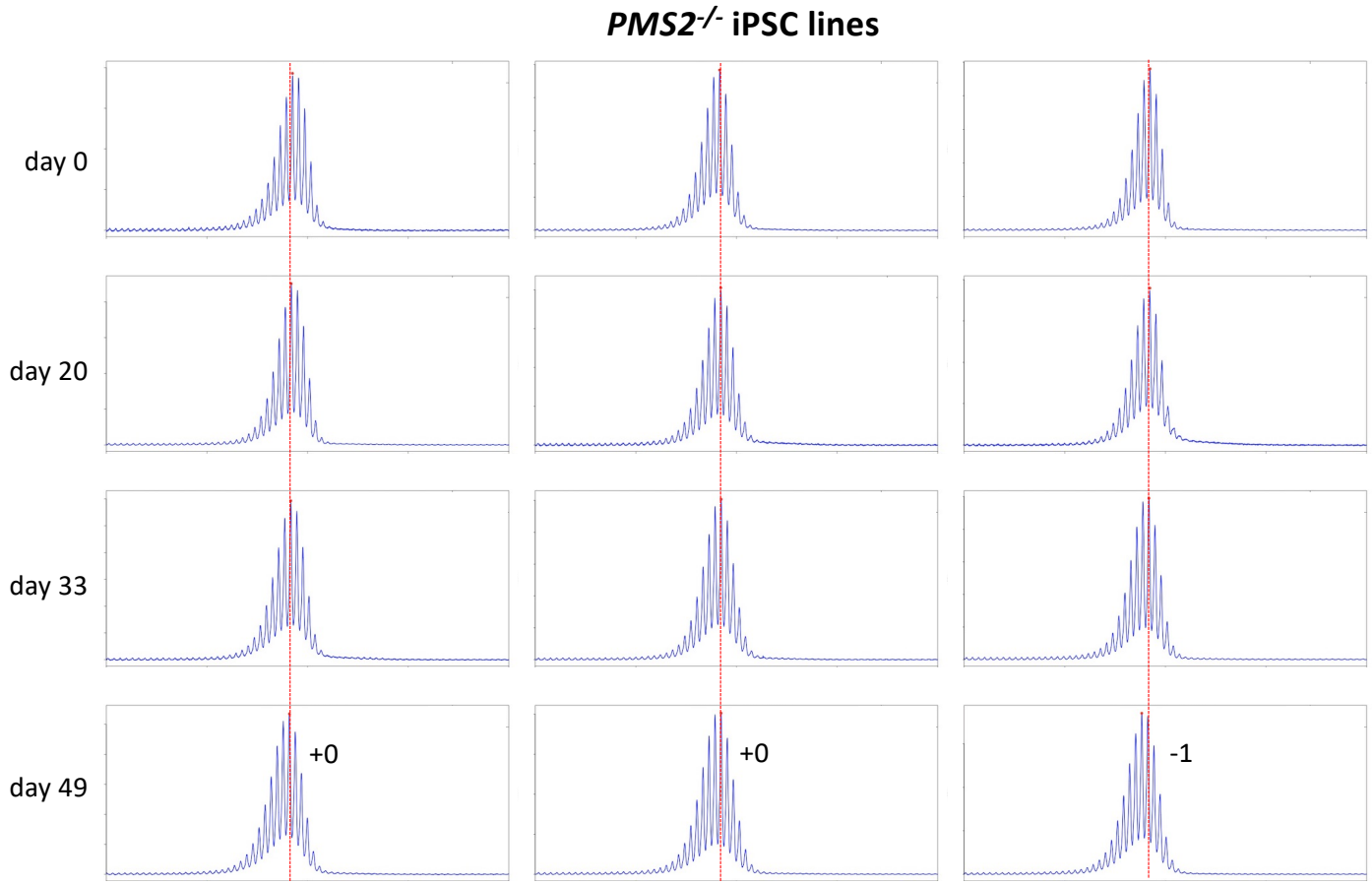

Fig. 5

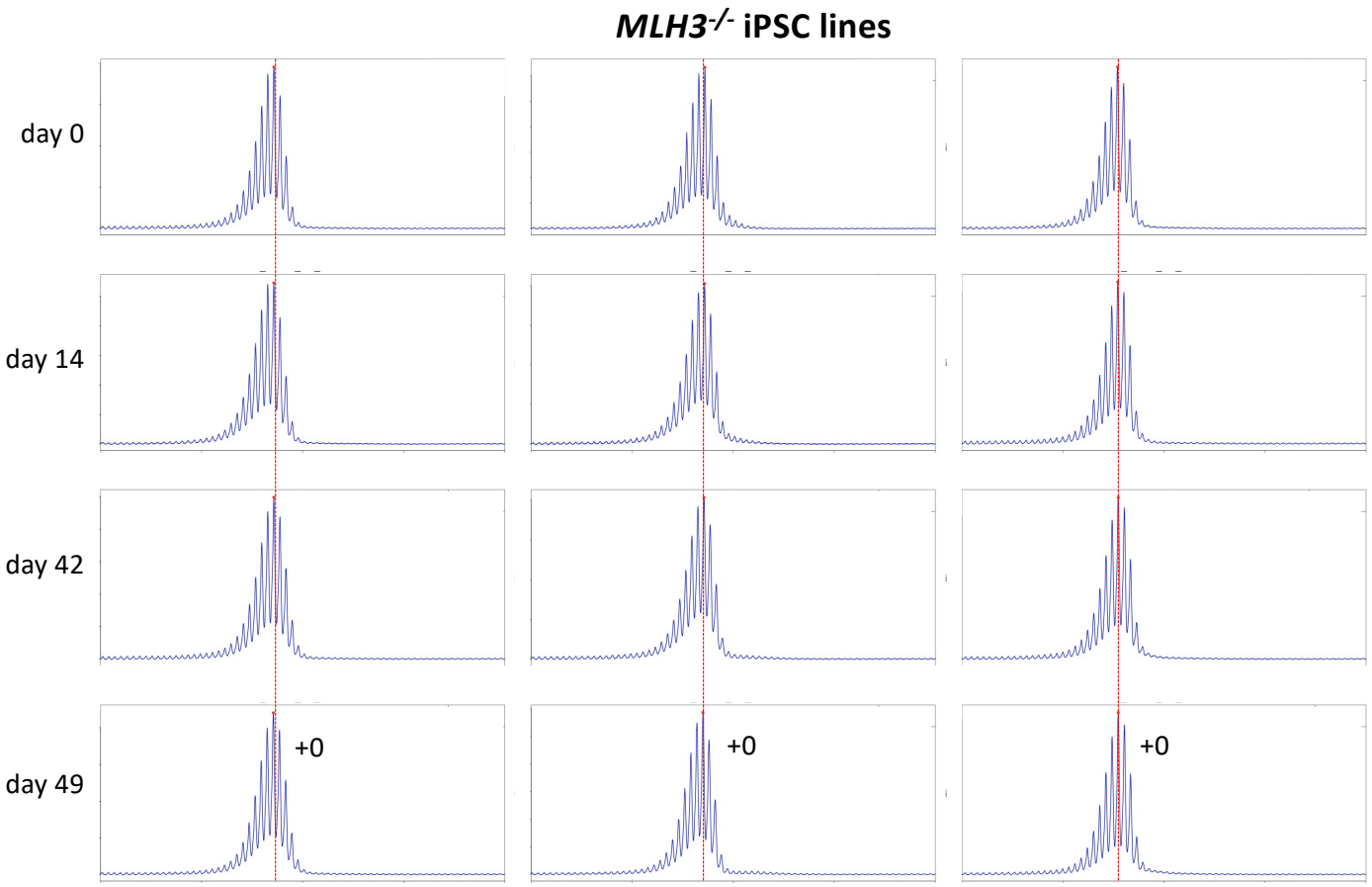

Fig. 6