Impeding transcription of expanded microsatellite repeats by deactivated Cas9

Belinda S. Pinto*1,3,4, Tanvi Saxena*1,3,4, Ruan Oliveira1,3,4, Héctor R. Méndez-Gómez1,4, John D. Cleary1,3,4, Lance T. Denes1,3,4, Ona McConnell1,3,4, Juan Arboleda1,3,4, Guangbin Xia2,3,4, Maurice S. Swanson1,3,4, and Eric T. Wang1,3,4

1Department of Molecular Genetics & Microbiology, 2Department of Neurology, 3Center for NeuroGenetics, 4University of Florida, Gainesville, Florida, 32610, USA

*These authors contributed equally to this publication.

Correspondence and Lead Contact: eric.t.wang@ufl.edu

Summary

Transcription of expanded microsatellite repeats is associated with multiple human diseases, including myotonic dystrophy, Fuchs’ endothelial corneal dystrophy, and C9orf72-ALS/FTD. Eliminating or reducing production of RNA and proteins arising from these expanded loci holds therapeutic benefit. Here, we tested the hypothesis that a deactivated form of the Cas9 enzyme impedes transcription across expanded microsatellites. We observed a repeat length-, PAM-, and strand-dependent reduction in the abundance of repeat-containing RNAs upon targeting dCas9 directly to repeat sequences. Aberrant splicing patterns were rescued in DM1 cells, and production of RAN peptides characteristic of DM1, DM2, and C9orf72-ALS/FTD cells was drastically decreased. Systemic delivery of dCas9/gRNA by adeno-associated virus led to reductions in pathological RNA foci, rescue of chloride channel 1 protein expression, and
decreased myotonia. These observations suggest that transcription of microsatellite repeat-containing RNAs is more sensitive to perturbation than transcription of other RNAs, indicating potentially viable strategies for therapeutic intervention.

**Introduction**

Microsatellite expansion diseases are a class of genetically inherited conditions associated with the destabilization and expansion of short repetitive sequence elements in the genome, which cause pathogenic effects via multiple mechanisms including epigenetic silencing, RNA gain of function, and/or protein gain of function\(^1\). These diseases, which include myotonic dystrophy Types 1 and 2 (DM1, DM2), Fuchs’ endothelial corneal dystrophy (FECD), Huntington Disease (HD), C9orf72-ALS/FTD (C9ALS/FTD) and spinocerebellar ataxias (SCAs), are often multi-systemic and can profoundly affect the central nervous system, muscle, and the heart. Somatic instability causes repeat expansion throughout the lifetime of an individual, with the most dramatic expansions reaching thousands of nucleotides in length in those post-mitotic tissues\(^2,3\). Cellular toxicity in these diseases partly occurs due to transcription of the expanded repeat tract. For example, in DM1 and DM2 expanded CUG or CCUG repeat RNAs, respectively, sequester Muscleblind-like (MBNL) proteins away from their endogenous RNA targets\(^4\) leading to aberrant splicing patterns\(^5\) and altered RNA stability/localization\(^6,7\), among other effects. In C9ALS/FTD, expanded G\(_4\)C\(_2\) and C\(_4\)G\(_2\) repeat RNAs sequester several RNA binding proteins, as well as undergo RAN translation to produce dipeptide polymers toxic to cells\(^8-10\). Taken together, these findings suggest that silencing of the expanded repeat locus holds therapeutic value.

Various approaches have been taken to silence toxic RNA or protein in microsatellite expansion diseases, including antisense oligonucleotides\(^5,11-13\), small RNAs\(^14-17\), and small molecules\(^18,19\). Perturbation to co-factors of RNA polymerase II (RNA Pol II) reduces transcription through expanded repeats in HD\(^20\) and C9ALS/FTD\(^21\) models, and treatment with Actinomycin D at nanomolar doses preferentially impedes transcription of CTG repeats in DM models\(^18\). A
prevailing hypothesis is that the efficiency of transcription through expanded repeats is decreased relative to non-repetitive sequences. This would provide a therapeutic window through which to impede transcription of these sequences in a repeat length-dependent manner resulting in premature termination and nascent transcript turnover.

A deactivated version of the Cas9 enzyme, of the clustered regularly interspaced short palindromic repeats (CRISPR) system, has been used to impair transcription of specific loci, as well as to visualize, tether, and/or isolate DNA in a sequence-specific manner. In prokaryotes, deactivated Cas9 (dCas9) can efficiently inhibit transcriptional initiation and elongation when bound to gene bodies or promoters\textsuperscript{22}. In eukaryotes, dCas9 can inhibit transcriptional initiation when fused to an inhibitory domain and targeted near the transcription initiation site\textsuperscript{23-26}. However, elongation inhibition by targeting dCas9 alone to the gene body has been largely ineffective, even when recruiting dCas9 to > 90 possible targeting sites\textsuperscript{27}.

Here, we test the hypothesis that expanded microsatellite repeats are highly sensitive to transcriptional blockade by dCas9, even in the context of processively elongating RNA Pol II. We find that the efficiency of inhibition follows rules similar to that observed in non-repetitive contexts, with clear dependencies on proto-spacer adjacent motif (PAM) sequence and the targeted DNA strand. In addition, application of this approach to cell and animal models of disease rescues downstream pathogenic consequences. These observations present a novel application of the CRISPR/Cas9 system, provide new tools to impede transcription of expanded microsatellites in a strand-dependent manner, and suggest that targeting transcription in a repeat length-dependent manner may be a viable therapeutic strategy for these diseases.

Results

dCas9-gRNA complexes reduce abundance of repeat-containing RNAs in a length-, PAM-, and strand-dependent manner
Binding of deactivated *S. pyogenes* Cas9 (dCas9) to DNA in the vicinity of transcriptional start sites (TSSs) can inhibit transcription in prokaryotes and eukaryotes when coupled with a transcriptional inhibitory domain. However, inhibition of transcriptional elongation via gene body targeting (>1kb away from TSS) has been relatively inefficient in eukaryotes. We hypothesized that transcriptional inhibition of expanded microsatellite repeats by dCas9 would be more potent as the repetitive nature of these repeats, sometimes thousands of tandem copies, would: 1) present challenges for RNA Pol II elongation even under normal conditions; and 2) allow for high levels of dCas9 recruitment using a single guide RNA (gRNA) sequence forming a substantial block to the elongating polymerase (Fig. 1A). To test this model, we employed a plasmid-based strategy in cell culture to examine the effects of dCas9 recruitment on transcription of CTG/CAG repeats that occur in DM1, FECD, HD, and SCAs. We utilized plasmids containing 0, 12, 40, 240, 480, or 960 CTG repeats located >1.5 kilobases downstream of the TSS in the *DMPK* 3’ UTR. To recruit dCas9 to these repeats, we designed gRNAs targeting CTG/CAG repeats in each of 3 possible nucleotide phases of the non-template strand, and 1 phase of the template strand (Fig 1B, left panel). Each CRISPR PAM was, therefore, constrained to CAG, AGC, GCA, or CTG. To precisely measure the expression of the repeat transcripts, we developed an amplicon-based deep sequencing assay called “Measurement of Barcoded Transcripts by Amplicon Sequencing” (MBTA-Seq) (Fig. S1). For this assay, we modified the plasmids by introducing distinct 8 nucleotide barcodes downstream of each CTG repeat tract. RNA was harvested from cells transfected with these plasmids, polyA+ selected, amplified by RT-PCR across the barcode region but avoiding the repeat containing region, and deep sequenced (Fig. S1). This assay was highly reproducible, and allowed us to simultaneously measure the expression of multiple repeat-containing transcripts in the same pool of cells (Table S1). In the presence of dCas9 and (CAG)$_6$ gRNA, we observed a dramatic reduction in expression of RNAs containing expanded CUG repeats (Fig. 1B, right panel). Knockdown efficiency increased as a function of the number of repeats present, presumably as a function of the number of dCas9-gRNA complexes that could be recruited. Twelve CTG repeats, likely recruiting at most a single dCas9-gRNA complex,
showed ~50% repeat-containing RNA remaining, and 40 CTG repeats showed ~25% remaining. Repeat lengths >240 CTG showed only ~5% remaining. (AGC)$_6$ and (GCA)$_6$ gRNAs showed poor knockdown efficiency, consistent with previously described SpCas9 PAM preferences, where NGG is best, NAG is second best, and NCG/NTG are equally disfavored$^{28}$. The (CUG)$_6$ gRNA showed little to no knockdown of transcripts containing expanded CUG repeats, potentially due to a weak PAM as well as targeting to the template strand. Importantly, dCas9 was required for decreased expression of repeat-containing transcripts, as presence of gRNAs alone did not lead to knockdown (Fig. S2A).

These observations suggest that recruitment of dCas9-gRNA complexes can impede transcription of expanded microsatellite repeat tracts in a repeat length-dependent manner. However, these experiments were performed using transiently transfected plasmids, which may not accurately model all aspects of transcriptional regulation in a genomic context. Furthermore, experiments to assess binding of protein complexes to DNA loci are most commonly performed using genomic targets and not plasmids. Therefore, we established a HeLa cell line in which transgenes for each of the 6 CTG repeat lengths, with associated barcodes, were stably integrated (Fig. S2B). To measure DNA binding by dCas9-gRNA complexes, we performed chromatin immunoprecipitation (ChIP) against dCas9 in the presence of (CAG)$_6$ gRNA or non-targeting control gRNA. We used deep sequencing of barcodes, similar to MBTA-Seq, to quantify abundance of immunoprecipitated DNA encoding each repeat length. To also confirm transcriptional repression in this cell line, we measured RNA abundance by MBTA-Seq following transfection of dCas9-gRNA complexes. We observed that relative dCas9-gRNA binding to the DNA increased as a function of repeat length, concomitant with a decrease in relative RNA abundance (Fig. 1C). In this genomic context, dCas9-gRNA reduced the abundance of transcripts containing 960 CUG repeats to ~20%, and was associated with ~8-fold increased binding relative to genomic loci with no CTG repeat tracts. These observations suggest that
binding of dCas9-gRNA complexes to DNA impedes transcription of repeat tracts, and that both binding and potency of transcriptional inhibition increases with the number of repeats.

To test the ability of dCas9-gRNA complexes to impede transcription of other repeat-containing sequences, we assayed CCTG repeat tracts, which cause DM2. dCas9-gRNA complexes similarly reduced expression of CCUG repeat-containing RNAs (Fig. 1D). Here, the tetranucleotide repeat allowed testing of four distinct gRNAs targeting the non-template strand, (CAGG)$_5$, (AGGC)$_5$, (GGCA)$_5$, and (GCAG)$_5$, corresponding to PAMs AGG, GGC, GCA, and CAG, respectively. Consistent with PAM preferences of SpCas9, knockdown was most efficient with the (CAGG)$_5$ gRNA, although the (AGGC)$_5$ and (GCAG)$_5$ gRNAs were also quite effective. The (CCUG)$_5$ gRNA was not effective at impeding transcription of CCTG repeats. Similar to our studies of expanded CUG repeats, presence of gRNAs alone did not lead to knockdown (Fig. S2C).

In both bacterial and mammalian systems, blockade of RNA Pol II has been shown to be most effective when targeting dCas9 to the non-template strand of transcribed DNA, likely because helicase activity native to RNA Pol II can unwind nucleic acids hybridized to the template strand$^{22}$. Our results from targeting repeats with (CAG)$_6$ gRNA versus (CUG)$_6$ gRNA and (CAGG)$_6$ gRNA versus (CCUG)$_6$ gRNA are consistent with this model, but these comparisons involve changes to both the targeted strand as well as the PAM. In addition, some of the most efficient gRNAs are those that could also facilitate dCas9 targeting of RNA via complementarity between the gRNA and the repeat-containing RNA. To study how the PAM sequence and targeted strand influence the efficacy of transcriptional blockade in isolation, and to clarify DNA versus RNA-targeting mechanisms, we generated (CAG)$_{960}$ and (CAGG)$_{240}$ plasmids suitable for MBTA-Seq. These constructs are identical to their (CTG)$_{960}$ and (CCTG)$_{240}$ repeat-containing counterparts except for their repeat tract and barcode sequences. We then used MBTA-Seq to measure the efficiency of CUG$_{960}$ or CAG$_{960}$ knockdown in the presence of either (CAG)$_6$ or (CUG)$_6$ gRNA (Fig. 1E). By
testing all four combinations, we could separate PAM-dependent effects from strand-dependent
effects. Consistent with previous reports, targeting the non-template DNA strand reduced
expression more effectively than targeting the template strand. Specifically, \((\text{CUG})_{960}\) was
reduced to \(~5\%\) by a \((\text{CAG})_6\) gRNA, while \((\text{CAG})_{960}\) was reduced to \(~40\%\) by the same gRNA.
These two conditions effectively compare targeting of non-template vs. template strand,
controlling for the CAG PAM. Similarly, \((\text{CAG})_{960}\) was reduced to \(~50\%\) by the \((\text{CUG})_6\) gRNA, and
\((\text{CUG})_{960}\) remained relatively high at \(~97\%\). Again, targeting of the non-template strand is more
efficient than targeting of the template strand, controlling for the CTG PAM.

PAM-dependent effects were quantified by comparing abundance of \((\text{CTG})_{960}\) in the presence of
\((\text{CAG})_6\) gRNA relative to abundance of \((\text{CAG})_{960}\) in the presence of \((\text{CUG})_6\) gRNA. Here, both
gRNAs target the non-template strand, but use different PAMs. We observed \(~5\%\) RNA
remaining with the CAG PAM, and \(~55\%\) RNA remaining with the CTG PAM, controlling for
targeted strand, consistent with the CAG PAM being more effective. Similarly, measurement of
\((\text{CAG})_{960}\) RNA in the presence of \((\text{CAG})_6\) gRNA and \((\text{CTG})_{960}\) RNA in the presence of \((\text{CUG})_6\)
gRNA allows comparison of both PAMs, controlling for targeted strand. Again, we observed more
effective silencing with a CAG PAM (\(~40\%\) RNA remaining) as compared to the CTG PAM (no
change in RNA). We observed similar trends with CCTG/CAGG repeats (Fig. 1F). \((\text{CCUG})_{240}\)
was reduced to \(~5\%\) by \((\text{CAGG})_5\) gRNA and \((\text{CAGG})_{240}\) to \(~55\%\) by \((\text{CAGG})_5\) gRNA (non-
template vs. template strand, AGG PAM). We did not observe the same template vs. non-
template effect when silencing with a CTG PAM, but the \((\text{CCTG})_5\) gRNA was largely ineffective,
showing little to no silencing.

Overall, these results separate the effects of PAM sequence and targeted strand in the context of
transcriptional blockade. Furthermore, they support a model in which dCas9/gRNA complexes
target repeat-containing DNA, because reductions in RNA abundance are achieved even when
using gRNAs that are not complementary to transcribed RNAs.
dCas9-mediated transcriptional inhibition rescues splicing defects and blocks RAN translation in cell-based models of DM and C9ALS/FTD

Downstream symptoms of many repeat expansion diseases are caused by the expression of toxic RNA species. In some cases, such as DM and ALS/FTD, the RNA sequesters RNA binding proteins necessary for cellular functions, leading to numerous downstream molecular changes to the transcriptome and proteome. To assess the extent to which dCas9-mediated transcriptional silencing can rescue molecular and cellular phenotypes in disease models, we used a HeLa cell-based model of DM1, in which (CUG)$_{480}$ repeats are expressed from a plasmid. Co-expression of dCas9 and (CAG)$_6$ gRNA led to a reduction in the percentage of cells showing CUG-containing RNA foci, as well as a reduction in the number of foci per cell (Fig. 2A, B). We measured the extent to which MBNL-dependent splicing misregulation, characteristic of DM1 cells, could be rescued using a splicing minigene reporter containing MBNL1 exon 5. The percent spliced in (psi, Ψ) of this exon is regulated by MBNL proteins, and changes from 10% in healthy HeLa cells to ~70% in cells expressing (CUG)$_{480}$ repeats. Co-transfection of dCas9 and (CAG)$_6$ gRNA partially rescued splicing dysregulation, reducing Ψ to ~35% (Fig. 2C). This rescue occurred only in the presence of both dCas9 and (CAG)$_6$ gRNA, indicating that (CAG)$_6$ gRNA expression alone cannot neutralize toxic CUG RNA by directly hybridizing to repeat RNA and displacing MBNL protein or facilitating degradation by dsRNA-sensitive nucleases. It is well established that splicing defects in DM1 depend on the extent of MBNL sequestration and therefore total CUG repeat load; Ψ values for MBNL1 exon 5 increased with the length of the CTG repeat tract transfected into cells (Fig. 2D). Interestingly, co-expression of dCas9 and (CAG)$_6$ gRNA yielded a similar rescue of splicing across all repeat lengths exhibiting mis-splicing, i.e. 240, 480, and 960 CUG repeats.

The CTG, CCTG, and G$_4$C$_2$ repeats associated with DM1, DM2, and C9ALS/FTD, respectively, undergo repeat-associated non-ATG (RAN) translation (Zu et al, submitted 2017). We
hypothesized that reduction in repeat-containing RNA abundance by dCas9-gRNA complexes would yield reduced RAN protein abundance. We assessed this by performing western analyses of RAN peptides robustly expressed from CTG, CCTG and G₄C₂ repeat containing plasmids transfected into HEK293T cells. Detection of the RAN products is facilitated by the presence of multiple tags downstream of the repeats, one in each coding frame (Fig. 2E). In all three western analyses, RAN proteins were observed only in the presence of repeat-containing plasmids, and were slightly reduced upon expression of a control, non-targeting gRNA, likely due to competition of plasmids for transcriptional machinery. Consistent with measurements of CUG-containing RNA levels (Fig. 1B), the FLAG-tagged poly-Ala peptide translated from CUG repeats was dramatically reduced in the presence of (CAG)_₆ gRNA, and modestly reduced in the presence of (CUG)_₆ gRNA (Fig. 2E). Importantly, the RAN protein was not reduced in the presence of (CAG)_₆ gRNA alone without dCas9 (Fig. S2D). HA-tagged poly-LPAC translated from CCUG repeats was dramatically reduced in the presence of (CAGG)_₅, (AGGC)_₅, and (GCAG)_₅ gRNAs (Fig. 2F), but not in the presence of (CCUG)_₅ gRNA (Fig. S2E). To test the effects of dCas9 on levels of poly-GlyPro translated from G₄C₂ repeats, we designed two gRNAs with NGG PAMs: (C₄G₂)_₃ that would target the non-template strand and (G₄C₂)_₃ that would target the template strand. Consistent with our studies of CTG and CCTG repeats, dCas9-gRNA complexes targeted to (G₄C₂)₁₂₀ exhibited silencing of RAN peptide in a strand-dependent manner (Fig. 2G), and did not occur in the presence of gRNA alone (Fig. S2F). While reduction in RAN protein was achieved with both gRNAs, silencing was more effective with gRNA targeting the non-template strand.

**dCas9-mediated transcriptional inhibition reduces nuclear RNA foci and rescues splicing defects in human DM1 myoblasts**

We next tested whether this approach could impede transcription of expanded repeats in the native DMPK locus, in primary human DM1 myoblasts. To express deactivated Cas9, we generated an adeno-associated virus (AAV) carrying sequences for the Cas9 and gRNAs that can efficiently infect muscle cells. As the traditional dCas9 sequence cannot be accommodated...
within the limits of the AAV genome, we turned to the smaller *S. aureus* Cas9 (SaCas9) that can be packaged with a U6 promoter-driven gRNA within the AAV genome\(^\text{35}\). We generated a deactivated SaCas9 (dSaCas9) by mutating catalytic residues D10 and H557 to alanines \(^\text{36}\). Since dSaCas9 exhibits slightly different PAM preferences to dSpCas9 (NNGRR versus NGG), we used MBTA-Seq to confirm that dSaCas9 with a (CAG)\(_6\) gRNA could impede transcription of expanded CTG repeats in a length-dependent manner (Fig. S3A). Next, we packaged dSaCas9 with control or (CAG)\(_6\) gRNA into an AAV2/6 virus, as we found that the AAV6 capsid efficiently infects myoblasts in culture (Fig. S3B).

We infected a primary DM1 myoblast line with the AAV2/6 encoding dSaCas9 and (CAG)\(_6\) or control gRNAs. Expression of dSaCas9 protein with 5 nuclear localization signals in the infected cells was detected by immunofluorescence against the HA tag and was found to be nuclear, although at times also appeared to be cytoplasmic (Fig. S3B, Fig. 3A). We assayed CUG repeat content by FISH and as observed previously in HeLa cells, the number of CUG repeat foci per cell was decreased in the presence of dSaCas9 and (CAG)\(_6\) gRNA, relative to control gRNA (Fig. 3A, B).

Although splicing events in DM1 patient muscle have been well characterized, myoblasts in culture do not express many transcripts typically present in mature muscle. To identify molecular changes in cultures that appropriately model *in vivo* DM1 biology, we further characterized the line described above, as well as a line derived from an individual unaffected by DM1. Four independent RNAseq libraries were created from the DM1 cells grown in several different growth conditions, and two from the unaffected line. To identify splicing changes in these lines that appropriately model those occurring in human DM1 muscle, we first defined a high confidence set of DM1-relevant splicing events by re-analyzing a set of 55 transcriptomes (44 DM1 and 11 unaffected) from human tibialis biopsies (GSE86356). We identified splicing events whose inclusion level, \(\Psi\), strongly correlated to the concentration of free, functional MBNL protein in the
affected tissue, as previously described\textsuperscript{34}. We fit sigmoid curves describing the relationship between MBNL concentration and $\Psi$, for example for MBNL1 exon 5 (Fig. 3C). The best-fitting events were selected (see Methods, Fig. S4), and their dysregulation ($\Delta\Psi$, Tibialis anterior, DM1 minus unaffected) was plotted against dysregulation observed in the DM1 myoblasts ($\Delta\Psi$, Primary myoblast, DM1 minus unaffected) (Fig. 3D). These filtering steps yielded a set of events whose behavior is reasonably recapitulated in each system; the correlation in splicing dysregulation between tibialis and myoblasts was $\sim$0.68.

To assess splicing rescue in DM1 myoblasts upon treatment with AAV-dSaCas9 and (CAG)$_6$ gRNA, we first performed RT-PCR to measure the inclusion level of MBNL1 exon 5. While viral infection with an AAV encoding dSaCas9 and control gRNA yielded a $\Psi$ of $\sim$50%, infection with AAV encoding dSaCas9 and (CAG)$_6$ gRNA rescued $\Psi$ to $\sim$20% (Fig. 3E). We assessed splicing transcriptome-wide by RNAseq, and analyzed those exons in the myoblasts whose changes are concordant with changes observed in tibialis biopsies (lower left and upper right quadrants, Fig. 3D), and whose baseline $\Psi$ values are less than 0.33 units apart when comparing tibialis to myoblasts. Several exons were successfully rescued, and a correlation of $\sim$0.66 was observed between dysregulation in tibialis ($\Delta\Psi$, DM1 minus unaffected) versus rescue by dSaCas9 and (CAG)$_6$ gRNA ($\Delta\Psi$, control gRNA minus (CAG)$_6$ gRNA) (Fig. 3F, Table S3). To assess potential off-target changes in gene expression for other transcripts with genomic CTG or CAG repeat tracts, we treated non-DM1 myoblasts with AAV-dSaCas9 and (CAG)$_6$ gRNA or control gRNA and performed RNA-Seq. We enumerated the longest contiguous CTG or CAG repeat tracts in all pre-mRNAs in the human genome, and found none to exceed 24 repeats; maximums were 24 CTG in TCF4 and 22 in AR. We observed no differences in gene expression that depended on the length of repeat tracts, for both CTG or CAG repeats (Fig. S3C, D).

These observations suggest that dSaCas9 targeted to CTG repeat tracts can impede transcription of expanded CUG repeats, relieve MBNL sequestration, and restore splicing
homeostasis in human cells containing DM1 repeat expansions, with selectivity that depends on the extreme repeat lengths commonly found in symptomatic tissue.

**dCas9-mediated transcriptional inhibition reduces RNA foci, rescues Clcn1 expression, and decreases myotonia in a mouse model of DM1**

We next assessed whether dCas9 could impede transcription of expanded CTG repeats in a well-established mouse model of DM1, *HSA*LR37. These mice carry a human skeletal actin transgene containing 250 CTG repeats in the 3' UTR and exhibit molecular, cellular, and phenotypic properties characteristic of DM patients. To assess the efficacy of this approach independently of potential *in vivo* drug delivery issues, we first dissected extensor digitorum longus muscle (EDL) fibers, and cultured them *ex vivo*. As CTG repeat expression in this model is driven by the *HSA* promoter, RNA foci are ubiquitous and numerous in myonuclei (Fig 4A, Top panel). To assess the extent to which dSaCas9 could reduce RNA foci, we infected cultured individual *HSA*LR EDL fibers for 2 days and performed RNA FISH against CUG repeats (Fig. 4A). In untreated *HSA*LR fibers, we observed robust FISH signal in ~80% of myonuclei. This remained at similar levels in the presence of a dSaCas9 with control gRNA, but was reduced to ~50% in the presence of dSaCas9 with (CAG)₆ gRNA, where any sign of FISH signal in a nucleus was recorded (Fig. 4B). Overall FISH signal intensity, quantitated in >1400 nuclei per condition (see Methods, Fig. S5), was dramatically decreased in fibers treated with dSaCas9-(CAG)₆ gRNA relative to control gRNA and untreated fibers (Fig. 4C). These observations suggest that efficient delivery of dSaCas9-(CAG)₆ to muscle fibers *ex vivo* is sufficient to significantly reduce RNA foci load within 2 days.

Having observed robust elimination of RNA foci *ex vivo*, we turned to *in vivo* experiments. Given previous reports of immune reactivity against Cas9, especially following intramuscular injection, we chose to administer AAV-dSaCas9-gRNA by temporal vein injection at postnatal day 2, prior to full establishment of immune tolerance. AAV2/6 or AAV2/9 carrying dSaCas9 and (CAG)₆ or
control gRNA was injected. Five weeks following injection, electromyography was performed to analyze myotonia in tibialis anterior and gastrocnemius muscles (Fig. 5A). Wild type FVB mice showed no myotonia, an Mbnl1 KO mouse showed myotonia in 100% of needle insertions, and HSA<sub>LR</sub> mice showed myotonia in 87-100% of insertions. HSA<sub>LR</sub> injected with AAV6-dSaCas9-control gRNA showed myotonia levels similar to uninjected HSA<sub>LR</sub> (87%). However, HSA<sub>LR</sub> injected with AAV6-dSaCas9-(CAG)<sub>6</sub> gRNA showed a reduction in myotonia, with some mice showing myotonia in 33-50% of insertions. Interestingly, nearly all mice injected with AAV2/6 showed rescue while only 1 mouse injected with AAV9 showed rescue (Fig. S5D).

Because mis-splicing of chloride channel 1 (Clcn1) is well-established to mediate myotonia in DM1 and in HSA<sub>LR</sub> mice, we measured splicing patterns of Clcn1 mRNA in treated mice. Surprisingly, when analyzing RNA extracted from bulk muscle tissue, we could not detect significant rescue in Clcn1 splicing, or other MBNL-dependent splicing events (data not shown). Because myotonia is measured by assaying individual bundles of fibers, we hypothesized that some regions of muscle may be rescued and not others, and that any potential change in isoform composition in these regions may be diluted in analyses of bulk tissue. This is consistent with a roughly ~2.5-fold increase in myonuclei per fiber over the first 4 weeks of post-natal development, which could also dilute the proportion of myonuclei containing AAV episomes<sup>39</sup>. AAV6-delivered dSaCas9 distribution in EDL fibers, as assessed by immunofluorescence against the HA tag, revealed mosaic expression, with rare fibers showing region-specific nuclear signal for dCas9 (Fig. 5B), and most fibers not showing presence of dCas9 at all. Control immunofluorescence experiments against lamin A showed ubiquitous and consistent labeling of all myonuclei, ruling out potential staining artifacts (Fig. S5E). In spite of mosaic dSaCas9 distribution, 5-15% of fibers showed complete loss of CUG RNA foci (Fig. 5C). Furthermore, animals that showed the greatest decrease in RNA foci also showed the strongest myotonia rescue. Reduction of RNA foci load in some nuclei raised the possibility that proper Clcn1 splicing could be achieved in a subset of nuclei, and that these mRNAs may spread locally throughout the fiber to produce functional
Clcn1 protein. Indeed, mice treated with (CAG)$_6$ gRNA showed increased Clcn1 staining at muscle membranes relative to mice treated with control gRNA, but Clcn1 protein was localized only to a subset of fibers per muscle section. These observations are consistent with observations of myotonia elimination in a subset of fibers, yet absence of splicing rescue when assessing bulk tissue (Fig. 5D). In summary, these experiments suggest that dCas9 can, in principle, rescue disease phenotypes via transcriptional repression, but that widespread rescue of molecular events in muscle will require efficient delivery to a large proportion of myonuclei expressing toxic RNA.

**Discussion**

The application of CRISPR/Cas9 to genetic diseases is thought to hold great therapeutic promise. Multiple groups have successfully edited genomes in cell and animal models of disease. Some have directly applied these approaches to microsatellite expansion diseases to remove the expanded repeat tract$^{40}$ or cause somatic instability leading to repeat contraction$^{41}$. However, challenges remain to ensure both efficient and specific cleavage activity across multiple cells and tissues of the body. Here, we explore an alternative form of Cas9, dCas9, in which cleavage activity is abrogated. Previous studies demonstrated that dCas9 can successfully reduce transcript levels in prokaryotes and this effect is achieved primarily by blocking elongation of RNA Pol II through the gene body rather than targeting the transcripts themselves$^{22}$. In eukaryotes, however, transcriptional inhibition by recruiting dCas9/gRNA complexes to the gene body was inefficient. As expanded microsatellites are thought to be more challenging for RNA Pol II elongation, we hypothesized that dCas9/gRNA recruitment would effectively impede elongating polymerase. By testing disease-associated repeat sequences across multiple repeat lengths in vitro and in disease models, we demonstrate that dCas9 can substantially reduce repeat-containing transcript abundance. Importantly, repression efficiency is proportional to repeat length, because longer repeats recruit an increased number of dCas9/gRNA complexes, leading to greater transcriptional blockade (Fig. 1B-C).
dCas9 is well established to target DNA\textsuperscript{22}, but has also been shown to bind repeat-containing RNA in a gRNA-dependent manner\textsuperscript{42}. Indeed, when we immunoprecipitated dCas9 and assessed RNA binding, we observed increased binding to transcripts containing longer repeats (Fig. S2G). However, as repeat length increases, the amount of successfully transcribed RNA decreases, suggesting that at long repeat lengths, less RNA remains to be targeted. In the context of disease, microsatellite repeat tracts undergo dramatic somatic expansion in DM1, DM2, C9ALS/FTD, and HD, and evidence suggests that repeat lengths in tissues of symptomatic individuals reach thousands of nucleotides in length\textsuperscript{2,3}. Therefore, DNA-based targeting by dCas9 may play a primary role in driving potential therapeutic benefits in individuals with full expansions (Fig. 6). At short and intermediate repeat lengths, therapeutic benefit may derive from both DNA- and RNA-based targeting by dCas9, to not only impede transcription of repeat-containing DNA, but also modulate downstream consequences of transcribed “escaper” RNAs. Taken together, we demonstrate that dCas9 is an effective tool to limit downstream effects of toxic RNAs and present proof of concept data suggesting that therapeutic benefit can be achieved through this approach.

MBTA-Seq allowed us to rapidly and accurately measure expression of transcripts that are traditionally difficult to assay, in particular across repeat types and repeat lengths in a highly controlled manner. In principle, this approach can be adapted for any collection of transcripts, in the context of transient transfections or stable cell lines. We used this assay to quantitatively measure how PAM sequence and target strand influence the efficacy of transcriptional blockade by dCas9; these effects likely directly relate to the strength of binding by Cas9 as well as the ability of RNA Pol II to unwind DNA. Although we show here that CTG, CCTG, and G\textsubscript{4}C\textsubscript{2} repeats are targeted effectively by dCas9, in principle, alternative Cas9 variants favoring PAM sequences encoded on the non-template strand of microsatellite repeats should also be effective. Importantly, by targeting gRNAs with similarly effective PAMs to template versus non-template strand transcripts, dCas9 can be used to separate effects arising from sense versus antisense
repeat-containing transcripts, such as those occurring in C9ALS/FTD.

Effective, safe delivery in a multi-systemic fashion to sufficient numbers of post-mitotic cells will be necessary for this approach to be therapeutically viable in the context of many of these dominantly inherited neurological diseases. In DM and C9ALS/FTD, it is unclear what fraction of nuclei must be rescued in multi-nucleated muscle or in the central nervous system to halt or improve disease symptoms. Here, we observed that rescue of a subset of nuclei was sufficient to yield partial rescue of myotonia. However, myotonia can be modeled as a loss-of-function event in which Clcn1 protein is lost, and there is evidence that in other recessive muscle diseases, gene editing of a small subset of nuclei can restore sufficient protein expression across muscle fibers to yield therapeutic benefit. Conversely, mis-splicing events yielding pathogenic isoforms with dominant-negative behavior may require full elimination to mitigate deleterious consequences.

Our approach used systemically delivered AAV, which revealed regional variation in transduction efficiency, likely because NLS-tagged Cas9 protein remained restricted to myonuclear domains transduced by AAV episomes, and many new myonuclei are recruited to fibers throughout the first 4 weeks of post-natal muscle development. It is possible that more localized or concentrated delivery would have facilitated transduction of a higher proportion of myonuclei. Previous studies with AAV-mediated gene therapy typically evaluate expression of proteins that can spread throughout muscle fibers, precluding measurement of transduction efficiency to all myonuclei. Regional variation in rescue is likely to be less relevant with therapies with more homogeneous delivery, for example, small molecules. Nevertheless, reduction of toxic RNA repeat load in a subset of nuclei was sufficient to increase production of Clcn1 protein, with consequences for distant regions of muscle naïve to dCas9. These results emphasize how differences in mechanism of action and delivery efficiency between various therapeutic approaches should significantly influence how we interpret changes in response to those therapies, both at the molecular and phenotypic level.
Finally, while this approach mitigates a key safety concern of CRISPR, which is the unwanted cleavage of off-targets, it also raises questions about whether long-term exposure to dCas9 can be achieved and tolerated, a necessity for repeat expansion diseases. Perhaps most importantly, the ability of dCas9 to impede transcription of microsatellite expansions defines a window in which transcriptional inhibition of expanded repeats is feasible, yet does not interfere with that of typical RNAs. This approach may serve as a baseline with which to compare alternative therapeutic approaches, as well as a tool to identify mechanisms and principles by which transcriptional activity of RNA polymerase II can be modulated in a sequence-specific manner.

Materials and Methods

Cell culture, cell lines and transfection

HeLa and HEK293T cells were cultured in 1X DMEM supplemented with 10% fetal bovine serum and 1% Penicillin/Streptomycin at 37°C and 5% CO₂. The DM1 primary myoblast cells were obtained by G. Xia. The clinical data of research subjects from who muscle biopsies were obtained, have been described before. Muscle biopsies were performed using 7G UCH Muscle Biopsy Needle (#8066, www.cadencescience.com). Samples were cut into small pieces and seeded into a 6cm dish. Myoblasts were expanded at a 1:2 ratio in myoblast growth media (Skeletal Muscle Cell Growth Medium-2, Lonza, #CC 3245) at 37°C and 5% CO₂. For viral transductions, these primary myoblasts were maintained in the growth medium for 3 days and then switched to Differentiation media (DMEM-F12 supplemented with 2% horse serum) for 3 days.

Plasmid DNA transfections of HeLa and HEK293T cells were performed using Trans-IT LT1 (MirusBio) as per manufacturer’s instructions. For RNA and FISH analyses, HeLa cells were harvested or processed 72 hours post transfection.
Cloning of barcoded repeat plasmids and gRNA constructs and generation of stable repeat expressing HeLa cell lines

Non-barcoded plasmids carrying 0, 12, 40, 240, 480, 960 CTG repeats (CTG\(_n\)) and 960 CAG repeats (CAG\(_{960}\)) were kind gifts from Tom Cooper (Baylor College of Medicine). Plasmids with 0 and 12 repeats were modified so that their vector backbones were identical to the others, and carried the ampicillin resistance gene. The (CCTG)\(_{240}\) plasmid was created as in Philips et al.\(^4\) In brief, oligonucleotide fragments 5'-TCGA(CCTG)\(_{20}\)C-3' and 5'-TCGAG(CAG)\(_{20}\)-3', were phosphorylated, annealed, gel isolated, and concatemerized by T4 DNA ligase. Concatemers not in a head to tail orientation were digested by SalI and XhoI. Concatemers were gel-isolated and cloned into the SalI site of CTG\(_0\). To create CAGG\(_{240}\), CTG\(_{960}\) was digested with SalI and HindIII and a new fragment was introduced which included HindIII and Agel restriction sites (DT_MCS). CCTG\(_{240}\) was digested with Xmal and HindIII and ligated with DT_MCS digested with HindIII and Agel to reverse orientation of the repeats and form CAGG\(_{240}\).

For MBTA-Seq, each repeat-containing plasmid was barcoded by introducing a fragment containing a random 8-nt sequence at the PflMI restriction site located downstream of the repeats via In-Fusion cloning (Clontech). Clones were sequenced to confirm that each repeat containing plasmid carried a unique barcode (Fig S1A). Barcoded CTG\(_{960}\) and CAG\(_{960}\) plasmids were generated by digesting (CTG)\(_{960}\) and (CAG)\(_{960}\) with HindIII and AccIII and ligating gBlocks containing unique barcodes with the digested plasmids.

To generate stable cell lines, the DMPK expression cassette was removed from each of the 6 barcoded plasmids and inserted into pAC156 (gift from Albert Cheng), a plasmid with Piggybac transposon terminal repeats as well as a puromycin selection cassette. All 6 plasmids were transiently transfected together with the Piggybac mPB transposase into HeLa, and selected by puromycin. Single cells were isolated by flow cytometry, and colonies were cultured in 96 well plates. 48 colonies were expanded and subjected to MBTA-Seq to screen for integration and
expression of all 6 plasmids. Counts for all barcode deep sequencing experiments can be found in Table S1.

The dCas9 gRNAs were cloned into an AflII-digested U6 expression vector by annealing oligos as previously described\textsuperscript{47}. The dSaCas9 gRNAs were cloned into BsaI-digested vector by annealing oligos as previously described\textsuperscript{35}.

**dCas9 chromatin and RNA IP**

dCas9 ChIP and RIP experiments were performed on the HeLa cell line containing the six different repeat lengths (CTG\textsuperscript{0} - 960) integrated into genome. 10cm plates of \~80% confluent cells were transfected with the pXIdCas9 and U6 expression vectors expressing the dCas9 protein and the control or (CAG)\textsubscript{6} gRNAs respectively\textsuperscript{48,47}, using the manufacturers guidelines. For ChIP, 48 hours after transfection cells were crosslinked with 1\% formaldehyde for 20 minutes at room temperature. Cross-linking reactions were stopped by addition of glycine to a final concentration of 0.125 M. Cells were then harvested, washed with phosphate buffered saline and pelleted. 1 mL of Lysis Buffer 1 (50mM Hepes [pH 7.5], 140mM NaCl, 1mM EDTA, 10\% glycerol, 0.5\% Igepal, 0.25\% Triton X-100) was added to the cells and rocked at 4\textdegree C for 10 minutes. After spinning, the cells were incubated in Lysis Buffer 2 (10mM Tris-HCl [pH 8.0], 200mM NaCl, 1mM EDTA, 0.5mM EGTA) for 10 minutes at RT. Nuclei were pelleted and resuspended in 1mL of Lysis Buffer 3 (10mM Tris-HCl [pH 8.0], 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.1\% Na-deoxycholate and 0.5\% N-lauroylsarcosine) and subjected to sonication in a Covaris S220 to obtain DNA fragments averaging 4kb in length. One-twentieth of the total chromatin served as input. The remaining material was used in the IP, which was performed using the HA-tag (C29F4) rabbit mAb conjugated to magnetic beads (Cell Signaling Technologies) at 4\textdegree C overnight to pull down the HA-tagged dSpCas9 and interacting DNA. Beads were washed 7 times with wash buffer (50mM Heps [pH 7.6], 500mM LiCl, 1mM EDTA, 1\% Igepal and 0.7\% Na-Deoxycholate). Immunocomplexes were eluted from the beads with elution buffer (50mM Tris-HCl, 10mM EDTA...
and 1% SDS) at 65°C for 15 minutes. Crosslinks in the IP and input were reversed overnight at 65°C and treated with RNase A and proteinase K to remove RNAs and proteins. DNA was extracted with phenol-chloroform and precipitated with ethanol. The barcoded region associated with each repeat length was amplified from the isolated DNA fragments using primers containing adapters facilitating deep-sequencing.

For the RNA IP experiments, 48 hours after transfection, cells were harvested in 1mL of lysis buffer (100mM Kcl, 5 mM MgCl2, 10mM HEPES [pH 7.0], 0.5% Igepal, 1mM DTT, 100 U/mL SUPERase In RNase inhibitor (Thermo Fisher), 2mM vanadyl ribonucleoside complexes solution, 25uL/mL protease inhibitor cocktail). One-twentieth of the resulting lysate was used as input and the remaining lysate was incubated with HA-tag (C29F4) rabbit mAb conjugated to magnetic beads (Cell Signaling Technologies) at 4°C overnight to pull down the HA-tagged dSpCas9 and interacting RNA. Beads were washed with the lysis buffer four times at 4°C and immunocomplexes were eluted off with 0.1% SDS and proteins removed using proteinase K at 50°C for 30 minutes. RNA was isolated using the Direct-zol RNA miniprep kit (Zymo Research) and contaminating DNA was eliminated using TURBO DNase (Thermo Fisher). cDNA was generated using Superscript IV Reverse Transcriptase (Thermo Fisher) and subsequently barcoded regions were amplified using flanking primers carrying sequences suitable for deep sequencing.

**FISH and immunofluorescence analyses**

To detect nuclear RNA foci, cells or muscle fibers were fixed with 4% PFA for 10 minutes at room temperature followed by ice cold RNAsase free 70% ethanol for 30 mins. Fixed samples were washed with a 25% formamide wash buffer at 30°C for 30 mins and then hybridized with a CalFluor 610 conjugated (CAG)10 oligonucleotide (Biosearch Technologies) in a 25% formamide hybridization buffer overnight at 30°C. Finally samples were washed two times at 30°C with wash buffer for 30 minutes to an hour, incubated with DAPI (1mg/mL) and mounted in Vectashield.
Further IF analysis was performed on HeLa cells and DM1 myoblasts to detect the presence of the dCas9-HA protein. After excess oligonucleotide was washed off, the cells were blocked in 3% normal goat serum in 1% Triton X-100-PBS for 30 mins at room temperature, incubated with anti-HA antibody (1:500, #3724, Cell Signaling Technologies) overnight at 4°C, washed with 1X PBS, incubated with Alexa Fluor 488 conjugated anti-rabbit secondary antibody (1:500, Life Technologies) for 2 hours, washed and incubated with DAPI for 5 mins and mounted in Vectashield. Slides were imaged using the Zeiss LSM 880 Confocal Laser Scanning Microscope.

IF analysis to detect dCas9 in mouse muscle fibers from the EDL of mice injected with AAV6-dSaCas9-(CAG)₆ was conducted as described above, except samples were fixed with 100% isopropanol at -20°C for 10 mins.

Clcn1 was detected in mouse muscle by performing IF on frozen muscle sections of the TA. Frozen muscle was sectioned into 10μM slices, fixed with 100% acetone at -20°C for 20 minutes, washed with 0.3% Triton X-100-PBS and incubated with rabbit anti Clcn1 (1:100, #CLC11-S, Alpha Diagnostic International) overnight at 4°C. Samples were incubated with goat anti-rabbit Alexa Fluor 568 (1:500, Thermo Fisher) for 2 hours at RT and then treated with DAPI and mounted in Vectashield.

**Western blots for RAN peptides**

HEK293T cells were transfected with one of the following plasmids that express tagged RAN translated products: pcDNA-6XStop-(CTG)₁₅₀-3X(FLAG-HA-cMyc-His), pcDNA-6XStop-(CCTG)₁₃₇-3X(FLAG-HA-cMyc-His) or pcDNA-6XStop-(G₄C₂)₁₂₀-3X(FLAG-HA-cMyc-His)⁹,¹⁰ (Zu et al, submitted 2017). These cells were co-transfected with the pXdCas9 plasmid expressing the dCas9 protein⁴⁸, and U6 expression vectors⁴⁷ expressing the control gRNA or gRNA’s targeting either strand of the CTG, CCTG and G₄C₂ repeats ((CAG)₆ or (CUG)₆ gRNAs, (CAGG)₅, (AGGC)₅, (GCAG)₅, or (CCUG)₅ gRNAs and (C₄G₂)₃ or (G₄C₂)₃ gRNAs, respectively).
72 hours after transfection, cells in each well of a 12-well tissue culture plate were gently rinsed 1X with PBS and lysed in 200ul of RIPA Buffer (50 mM Tris-Cl pH 7.4, 150 mM NaCl, 0.1% Na-Deoxycholate, 1% NP-40, 0.5% SDS) with protease inhibitors for 30 minutes on ice. Genomic DNA was sheared by 8-10 passages through a 21-gauge needle. The resulting lysate was centrifuged at 18,000 x g for 15 min and the supernatant was collected. The protein concentration of the lysate was determined using Pierce™ BCA Protein Assay. Equal amounts of protein were loaded and separated on a 4-12% NuPage Bis-Tris gel (Novex) and transferred to a nitrocellulose membrane (Amersham). The membrane was blocked in 5% milk in PBS-Tween20 (0.05%) for 1 hour and probed with anti-FLAG (1:2000) or anti-HA (1:1000) antibody in 1% milk solution in PBS-Tween20 (0.05%) overnight at 4°C. After the membrane was incubated with anti-mouse and anti-rabbit HRP (1:10,000) for 2 hours at room temperature, the bands were detected using the SuperSignal™ West Femto Maximum Sensitivity Substrate as per manufacturers protocol ⁹.

**Recombinant AAV production**

Viral production was achieved through transfection of HEK293T cells cultured in 150mm plates with the pAAV6 serotype packaging plasmid⁴⁹, pXX6 helper plasmid that contains the adenovirus E4, VA and E2a helper regions⁵⁰ and AAV2-ITR containing plasmid expressing dSaCas9 and the control or (CAG)₆ gRNA (generated from the SaCas9 plasmid pX601-AAV-CMV::NLS-SaCas9-NLS-3xHA-bGHpA;U6::Bsal-sgRNA, Addgene #61591). Transfections were carried out using the TransIT-LT1 transfection reagent and recommended protocols (MirusBio). Cells were harvested between 48h and 72h post-transfection, recombinant AAV2/6 virus was purified by iodixanol step gradients followed by vector concentration and buffer exchange with lactated Ringer’s in an Apollo 150kDa concentrator (Orbital Biosciences)⁵¹. Virus titers were determined using the Quant-iT Picogreen dsDNA assay kit (Life Technologies)⁵² and found to be ~10¹¹ vg/mL.

**Isolation of single muscle fibers**
Single muscle fibers were isolated from 3-4 week old \(HSA^{LR}\) mice as described previously\textsuperscript{53}. Briefly, the EDL was dissected and digested with a 0.2% Collagenase Type I in DMEM solution in a 37°C water bath for 1 hour without agitation. The digested muscle was then flushed with DMEM to separate out individual muscle fibers. Fibers were cultured in DMEM containing 20% FBS overnight before infection with AAV.

**Transduction of human DM1 myoblasts and isolated \(HSA^{LR}\) muscle fibers**

Virus carrying dSaCas9 and control or \((CAG)_6\) gRNA was used to infect human DM1 primary myoblast cell lines and \(HSA^{LR}\) mouse EDL muscle fibers. To determine whether blocking expression of the CTG repeats in the human myoblasts affected the presence of RNA foci and splicing of MBNL targets, cells were grown to 60% confluence on CC2 chamber slides and infected for 6 days (3 days in growth media plus 3 days in differentiation media) with viral titers of \(10^9\). To analyze the effects on RNA foci in \(HSA^{LR}\) muscle fibers, 10-20 muscle fibers were cultured in wells of a 96 well plate and infected with a \(10^9\) viral titer for 48h.

**Quantitation of Signal Intensity of Nuclear Foci in Mouse Muscle Fibers**

Python scripts were written to quantitate intensity of FISH signal from RNA foci within nuclei of muscle fibers (python functions are listed below, and also see Fig. S5). Regions of interest were defined across each tile-scanned z-stack image obtained by confocal imaging. To identify and segment nuclei, 16-bit intensity values were scaled to lie between 0 and 1 \(\text{skimage.img_as_float}\). Then, a gaussian filter of standard deviation 3 pixels was applied to the raw DAPI signal \(\text{skimage.filters.gaussian}\), and a threshold of mean + 5 * standard deviation of the gaussian filtered signal was used to generate a binary mask. Holes were filled \(\text{scipy.ndimage.fill_holes}\), and a binary opening operation was performed to remove salt noise \(\text{skimage.morphology.binaryOpening with ball structured element of radius 2}\). Nuclei were segmented and labeled \(\text{skimage.measure.label}\), and objects with pixel volume <1000 or >20000 were removed. FISH signal was scaled to lie between 0 and 1 as above, and a grayscale opening
operation was performed to measure background (skimage.morphology.opening with a ball structured element of radius 3). This background intensity multiplied by 3 was subtracted from the FISH signal to yield background-subtracted FISH signal. The binary nucleus mask was applied to this signal, and total intensity was measured within each nucleus, and divided by the nuclear volume in pixels, to obtain the final FISH signal per unit volume for each nucleus. This procedure was applied to all regions of interest across all fibers.

Analysis of RNAseq data

100 ng of RNA was used to prepare RNA-Seq libraries using the KAPA Ribo-Erase Strand-Specific kit. Samples were pooled and sequenced on the NextSeq 500 Version 2, using a High-Output 2x75 kit. Reads were mapped to hg19 by Hisat2, and splicing events were quantitated by MISO. \( \Psi \) values from DM tibialis biopsies were fit to sigmoid curves using 4-parameter estimation, where \( \Psi = \Psi_{\text{min}} + (\Psi_{\text{max}} - \Psi_{\text{min}}) / (1 + e^{-\text{slope} \times ([\text{MBNL}]_{\text{inferred}} - EC50)}) \), using python/scipy packages. The \([\text{MBNL}]_{\text{inferred}}\) value was taken from Wagner et al\(^{34}34\). The “fit error” was evaluated by taking the sum of squared errors between observed \( \Psi \) and \( \Psi \) as predicted by the sigmoid curves. Events consistently regulated between non-DM1 and DM1 myoblasts were identified using a modified monotonicity test\(^5454\), \( \Delta \Psi > 0.1, \ BF > 5 \) where the 2 non-DM1 libraries were grouped together, and 4 DM1 libraries were grouped together. For Figure 5B, events with <1.3 sigmoid fit error and >1 monotonicity Z-score were selected for display. For Figure 5F, only events identified in Figure 5B to lie in the upper right or lower left quadrants were further analyzed; in addition, events were required to exhibit <0.33 difference in \( \Psi \) between cells treated with AAV-dSaCas9-control gRNA and non-DM1 tibialis biopsies. Raw RNA-Seq reads for these libraries are publicly available (GEO accession number pending). Statistics on read coverage are in Table S2.

Electromyography
To determine whether expression of dCas9-(CAG)$_6$ rescued myotonia in the HSA$^{LR}$ mice, mice were injected with AAV6-dSaCas9 and control or (CAG)$_6$ gRNA at $10^{10}$ viral genomes per mouse via the temporal vein at P2$^{55}$. Myotonia was assessed by electromyography at 5 weeks of age. Electromyography was performed under general anaesthesia (intraperitoneal ketamine, 100 mg/kg; xylazine, 10 mg/kg) using 30 gauge concentric needle electrodes to examine the hindlimb gastrocnemius and tibialis anterior muscles. At least 15 needle insertions were performed in each muscle and myotonic discharges were denoted as a percentage of the total number of insertions.

**Author Contributions**


**Acknowledgements**

We would like to thank L. Ranum, T. Zu and J. Cleary for generously providing us with the plasmids to assess RAN translation, G. Aslanidi for providing the AAV6 capsid plasmid and J. Singh for advice on AAV production. We thank T. Wang for construction of the CCTG$_{240}$ repeat tract. We thank C. Thornton for RNA samples from DM1 patients, which were obtained from the BioBank of the University of Rochester Wellstone Center (NIH NS048843). E.T.W. and T.S. were supported by DP5 OD017865. L. D. was supported by a Thomas H. Maren Fellowship.

**Figure legends**

Figure 1. Deactivated SpCas9 impedes transcription of expanded microsatellite repeats in a length-, PAM-, and strand-dependent manner. A) Proposed model for how recruitment of
multiple dCas9/gRNA complexes to expanded microsatellite repeats impedes transcription by RNA Pol II. B) Schematic of gRNAs used to target transcription of CTG repeats (left panel). Abundance of CUG repeat-containing RNA in the presence of dCas9/gRNAs targeting the repeat tracts of various lengths in HeLa cells, relative to the same RNA species with zero repeats (right panel). Error bars show standard deviation. C) Relative dCas9 ChIP signal across all repeat lengths versus percent RNA remaining following (CAG)$_6$ gRNA treatment relative to control gRNA treatment. dCas9 ChIP signal is computed as dCas9 immunoprecipitation divided by input chromatin in the presence of (CAG)$_6$ gRNA, divided by dCas9 immunoprecipitation divided by input chromatin in the presence of control gRNA. Relative abundance of repeat-containing loci and RNAs was assessed by deep sequencing of associated barcodes. Error bars show standard error of the mean. D) Abundance of CCUG repeat-containing RNA in the presence of dCas9 and various gRNAs targeting the (CCTG)$_{240}$ repeat tract in HeLa cells, relative to the same RNA species with zero repeats. E) Abundance of RNAs containing 960 CUG repeats or 960 CAG repeats in the presence of dCas9 and (CUG)$_6$ or (CAG)$_6$ gRNAs in HeLa cells, relative to RNA species with zero repeats. Arrows denote comparisons relevant for assessing PAM- or strand-dependent effects on efficacy. F) Abundance of RNAs containing 240 CCUG repeats or 240 CAGG repeats in the presence of dCas9 and (CCUG)$_5$ or (CAGG)$_5$ gRNAs in HeLa cells, relative to RNA species with zero repeats. Arrows denote comparisons relevant for assessing PAM- or strand-dependent effects on efficacy. (two-tailed T-test, **p < 0.0005, *p < 0.005)

**Figure 2. Transcriptional inhibition rescues molecular and cellular phenotypes in cell culture models of DM1, DM2, and C9ORF72/ALS/FTD.** A) Representative images of HeLa cells transfected with plasmids expressing (CTG)$_{480}$ repeats, dCas9 tagged with HA and control or (CAG)$_6$ gRNAs. RNA foci are visualized by a (CAG)$_{10}$ fluorescent oligonucleotide (red), dCas9 by an α-HA antibody (green), and DNA by DAPI (blue). Scale bar: 10µm B) Quantitation of the number of HA-positive cells showing nuclear RNA foci (left panel) and the cumulative density function of HA-positive cells with a given number of RNA foci (right panel), in the presence of
dCas9 and control or (CAG)$_6$ gRNAs (Kolmogorov–Smirnov test, *p < 0.005). C) Percent spliced in (Psi, Ψ) of MBNL1 exon 5 expressed in a minigene context in HeLa cells, in the presence of combinations of plasmids encoding (CTG)$_{480}$ repeats, dCas9, and (CAG)$_6$ or (CUG)$_6$ gRNAs. N > 3 for all conditions (two-tailed T-test, *p < 0.005). D) Ψ of MBNL1 exon 5 expressed in a minigene context in HeLa cells, in the presence of plasmids encoding 0, 12, 40, 240, 480, or 960 CTG repeats, as well as dCas9 and (CAG)$_6$ or (CUG)$_6$ gRNAs. N > 3 for all conditions. Error bars are too small to be visible in this plot. E) Western blot against the FLAG-tagged poly-Ala RAN product expressed from DM1 CTG repeats transfected into HEK293T cells. Cells were transfected with various combinations of plasmids encoding no repeats or (CTG)$_{105}$ repeats with combinations of dCas9 and control, (CAG)$_6$ or (CUG)$_6$ gRNAs. The RAN peptide migrates as a smear between 75 and 150 kD. The asterisk indicates protein produced by the dSpCas9 plasmid, which cross-reacts with the anti-FLAG antibody. F) Western blot against the HA-tagged LPAC RAN protein expressed from DM2 CCTG repeats transfected into HEK293T cells. Cells were transfected with various combinations of plasmids encoding no repeats or (CCTG)$_{137}$ repeats with combinations of dCas9 and control, (CAGG)$_5$, (AGGC)$_5$, (GCAG)$_5$ or (CCUG)$_5$, gRNAs. The RAN peptide migrates around 20 kD. G) Western blot against the FLAG-tagged Poly-GlyPro RAN protein expressed from ALS G$_4$C$_2$ repeats transfected into HEK293T cells. Cells were transfected with various combinations of plasmids encoding no repeats or (G$_4$C$_2$)$_{120}$ repeats with combinations of dCas9 and control, (C$_4$G$_2$)$_3$ or (G$_4$C$_2$)$_3$ gRNAs. The RAN peptide migrates as a smear between 75 and 120 kD. The asterisk indicates protein produced by the dSpCas9 plasmid, which cross-reacts with the anti-FLAG antibody. β-Actin serves as the loading control for A, B and C. N = 3 for A, B and C (representative blot shown).

Figure 3. AAV-dSaCas9 rescues molecular and cellular phenotypes in human DM1 myoblasts. A) Representative images of DM1 myoblasts infected with AAV-dSaCas9 carrying control or (CAG)$_6$ gRNAs. RNA foci are visualized by a (CAG)$_{10}$ fluorescent oligonucleotide (red), HA-tagged dCas9 by an α-HA antibody (green), and DNA by DAPI (blue). Scale bar: 10µm B)
Probability density function of cells with a given number of RNA foci, in the presence of control or (CAG)$_6$ gRNAs. A Kolmogorov–Smirnov test was performed for statistical significance. C) MISO estimates of MBNL1 exon 5 $\Psi$ in 11 unaffected and 44 DM1 tibialis anterior biopsies are plotted in order of [MBNL]$_{\text{inferred}}$ as previously described. A sigmoid curve was fit to these points and shown (dashed line) with 90% confidence intervals (gray shading). MISO estimates of MBNL1 exon 5 $\Psi$ values in an unaffected and DM1 primary myoblast line (2 and 4 replicates, respectively) are shown in blue and red lines, respectively, with ranges also shaded. D) Scatter plot for splicing events regulated in both tibialis biopsies and the human myoblast lines described in (C). 115 splicing events were selected from tibialis biopsies based on best sigmoid fits with [MBNL]$_{\text{inferred}}$ as in (C). The x-axis of the scatter plot is the mean $\Psi$ in the most severely affected biopsies (<0.33 [MBNL]$_{\text{inferred}}$) minus the mean $\Psi$ across unaffected individuals. The y-axis of the scatter plot is the mean $\Psi$ in the DM1 myoblast minus the mean $\Psi$ in the unaffected myoblast line. Labeled points have $\Delta\Psi$>0.3 in both conditions, a sigmoid fit<1.3, and a y-axis monotonicity score>1 (see Methods). Pearson correlation for all points is listed. E) MBNL1 exon 5 $\Psi$, assessed by RT-PCR in unaffected myoblasts and DM1 myoblasts infected with AAV-dSaCas9 carrying control or (CAG)$_6$ gRNAs. F) Scatter plot illustrating changes in $\Psi$ in response to AAV-dSaCas9 (CAG)$_6$ gRNA, for splicing events exhibiting concordant behavior in (D). The x-axis is as in (D), but the y-axis is $\Psi$ as assessed in the DM1 myoblast minus the mean $\Psi$ in the unaffected myoblast line. Labeled points have $\Delta\Psi$>0.1 in both conditions and a y-axis Bayes Factor>5 (see Methods). Pearson correlation for all points is listed.

Figure 4. AAV-dSaCas9 reduces RNA foci in $HSA^{LR}$ mouse muscle fibers. A) FISH to detect nuclear RNA foci by a fluorescent oligonucleotide (CAG)$_{10}$ probe (magenta) in the myonuclei of EDL muscle fibers from the $HSA^{LR}$ mouse that were untreated (top), infected with AAV-dSaCas9-control gRNA (middle) or AAV-dSaCas9-(CAG)$_6$ gRNA (bottom). One representative fiber for each condition is shown on the left (Scale bar: 100$\mu$m) and insets from each fiber (white boxes) are shown on the right (Scale bar: 50$\mu$m). DAPI is in cyan. B) The percentage of myonuclei per
fiber showing RNA foci was quantitated across all 3 conditions with 5-6 fibers per condition. Each fiber contains 400-500 nuclei in all cases (two-tailed T-test, **p < 0.0005). C) Probability density function of intensity of FISH signal in myonuclei from untreated fibers (red), fibers infected with AAV-dSaCas9-control gRNA (orange) or AAV-dSaCas9-(CAG)_6 gRNA (yellow). A Kolmogorov–Smirnov test was performed for statistical significance.

Figure 5. AAV6-dSaCas9-(CAG)_6 rescues muscle phenotypes in HSA^{LR} mice. A) Percent needle insertions showing a myotonic run in wild type FVB, MBNL1^-/-, HSA^{LR}, (black circles) and HSA^{LR} treated with AAV6-dSaCas9 with control gRNA (orange circles) or (CAG)_6 gRNA (yellow circles). Each point represents a muscle from a single animal, but in some animals, both tibialis anterior and gastrocnemius were assayed. N=2 mice for control gRNA and N=4 mice for (CAG)_6 gRNA. B) Immunofluorescence against HA-dSaCas9 in a single EDL fiber from a mouse treated with AAV2/6-dSaCas9 and (CAG)_6 gRNA. DAPI is also shown in the merged image. Scale bar, 50 µM. C) Quantitation of numbers of EDL fibers showing complete, partial, or no CUG RNA foci upon treatment with AAV2/6-dSaCas9 and control gRNA or (CAG)_6 gRNA. Myotonia levels and numbers of fibers analyzed are also listed. D) Immunofluorescence against Clcn1 in tibialis muscle sections from WT FVB, HSA^{LR}, and HSA^{LR} treated with AAV-dSaCas9 and control gRNA or (CAG)_6 gRNA. DAPI is shown in blue. Scale bar, 50 µM.

Figure 6. Model for microsatellite repeat expansion targeting by dCas9. dCas9 binding to DNA impedes transcription of long, expanded microsatellite repeats. In the context of shorter repeats, fewer copies of DNA-bound dCas9 may be insufficient to fully inhibit elongation by RNA polymerase II, permitting production of RNA that can also be targeted by dCas9.

References


Figure 1

**A**
Productive transcription

Transcription blocked by CRISPRi

**B**
gRNAs

(CAG)$_6$

(AGC)$_6$

(GCA)$_6$

\[ \text{dCas9/gRNA complexes} \]

**C**

(CUG)$_n$ expression, (CAG)$_6$ gRNA relative to control

\[ \text{dCas9 ChIP, (CAG)$_6$ gRNA relative to control} \]

**D**

(CUG)$_{\text{exp}}$ expression relative to control

Targeted PAM

AGG GGC GCA CAG CTG

**E**

Expression of repeat relative to control

Transcribed repeat

Targeted PAM

PAM effect

Targeted strand

Strand effect

**F**

Expression of repeat relative to control

Transcribed repeat

Targeted PAM

PAM effect

Targeted strand

Strand effect
Figure 2

A

Control gRNA  (CAG)_6 gRNA

Merge (CAG)_10 DAPI

(CAG)_6

(CUG)_6

α-HA

B

Percent cells containing RNA foci

Cumulative Frequency

Number of foci per cell

C

Psi, Mbnl1 exon 5

(CUG)_480 dCas9 (CAG)_6 gRNA (CUG)_6 gRNA

D

Number of CTG repeats

Control gRNA  gRNA(CAG)_6

E

pcDNA3.1-(CTG)_105

dSpCas9

No transfection  No repeats  No gRNA  Control gRNA (CAG)_6 gRNA (CUG)_6 gRNA

kDa

α-FLAG (polyA)

150 75 42

α-Actin

F

pcDNA3.1-(CCTG)_137

dSpCas9

No transfection  No repeats  No gRNA  Control gRNA (CAG)_6 gRNA (AGGC)_5 gRNA (GCAG)_5 gRNA (CUG)_6 gRNA

kDa

α-HA (polyLPAC)

20

α-Actin

G

pcDNA3.1-(C4G2)_120

dSpCas9

No transfection  No repeats  No gRNA  Control gRNA (CAGG)_5 gRNA (CCUG)_5 gRNA (GCAG)_5 gRNA (CCUG)_5 gRNA

kDa

α-FLAG (polyGP)

20

α-Actin

*
Figure 3

A

Merge (CAG)_{10} DAPI (CAG)_{10} α-HA
dSaCas9 ctrl

dSaCas9 CAG

B

Uninfected (210 cells) [**] Uninfected (160 cells) [***] Uninfected (194 cells)

B

Number of foci per cell

C

DM1 myoblasts (n=4)

Unaffected myoblasts (n=2)

ψ, MBNL1 exon 5

[MBNL]_{inferred}

0.00 0.25 0.50 0.75 1.00

0.25 0.50 0.75 1.00

D

Δψ, Primary myoblasts (DM1 - Unaffected)

Δψ, Tibialis Anterior (DM1 - Unaffected)

Δψ, Mbnl1 exon 5

0.00 0.25 0.50 0.75 1.00

0.25 0.50 0.75 1.00

E

F

ψ, Mbnl1 exon 5

0 20 40 60 80 100

0 20 40 60 80 100

DM1 myoblasts

Uninfected (210 cells)
dSaCas9 ctrl (160 cells)
dSaCas9 CAG (194 cells)

ψ, Mbnl1 exon 5

ψ, DM1 myoblasts (dSaCas9_{CAG}) - dSaCas9_{ctrl}
Figure 4

A

(CAG)$_{10}$

Uninfected

dSaCas9

dSaCas9$_{CAG}$

B

Percent myonuclei with CUG foci (per fiber)

Uninfected (2014 cells) dSaCas9$_{ctrl}$ (1433 cells) dSaCas9$_{CAG}$ (1924 cells)

C

Mean intensity per unit volume in nucleus (a. u.)

Uninfected (2014 cells) dSaCas9$_{ctrl}$ (1433 cells) dSaCas9$_{CAG}$ (1924 cells)

*p < 10$^{-93}$  **p < 10$^{-116}$
Figure 5

Panel A: Percent needle insertions showing myotonia.

- WT
- MBNL1^+/−
- HSA^LR
- Control gRNA
- (CAG)^6 gRNA

Panel B: Immunostaining with α-HA and DAPI.

Panel C: Percent EDL fibers

- RNA Foci: Complete
- RNA Foci: Partial
- RNA Foci: None

Legend:
- Orange: Control gRNA (87% myotonia, 32 fibers)
- Yellow: (CAG)^6 gRNA (50% myotonia, 34 fibers)
- Light yellow: (CAG)^6 gRNA (33% myotonia, 69 fibers)

Panel D: Immunostaining with α-Clcn1 and DAPI.

- WT
- HSA^LR
- Control gRNA
- (CAG)^6 gRNA
Figure 6

- Short Repeats
  - Microsatellite repeat length
  - Toxic RNA levels
  - DNA-bound dCas9
  - RNA-bound dCas9

- Long Repeats