CELLULAR NEUROSCIENCE

Three-dimensional chromatin interactions remain stable upon CAG/CTG repeat expansion

Gustavo A. Ruiz Buendía[1], Marion Leleu[2,3], Flavia Marzetta[3], Ludovica Vanzan[4], Jennifer Y. Tan[5], Victor Ythier[4], Emma L. Randall[6], Ana C. Marques[5], Tuncay Baubec[7], Rabih Murr[4,8], Ioannis Xenarios[1], Vincent Dion[6]*

Expanded CAG/CTG repeats underlie 13 neurological disorders, including myotonic dystrophy type 1 (DM1) and Huntington's disease (HD). Upon expansion, disease loci acquire heterochromatic characteristics, which may provoke changes to chromatin conformation and thereby affect both gene expression and repeat instability. Here, we tested this hypothesis by performing 4C sequencing at the *DMPK* and *HTT* loci from DM1 and HD–derived cells. We find that allele sizes ranging from 15 to 1700 repeats displayed similar chromatin interaction profiles. This was true for both loci and for alleles with different DNA methylation levels and CTCF binding. Moreover, the ectopic insertion of an expanded CAG repeat tract did not change the conformation of the surrounding chromatin. We conclude that CAG/CTG repeat expansions are not enough to alter chromatin conformation in cis. Therefore, it is unlikely that changes in chromatin interactions drive repeat instability or changes in gene expression in these disorders.

INTRODUCTION

The genome is organized into hierarchical chromatin contact domains (*1*). This three-dimensional (3D) organization of chromatin in the nucleus has a profound impact on transcription, DNA replication, recombination, and repair (*1*, *2*). For instance, heterochromatic and euchromatic loci are spatially separated and display distinct 3D chromatin interactions as gleaned by microscopy and chromosome conformation capture (3C)–based experiments (*2*, *3*). However, how this higher-order chromatin structure impinges on biological functions, what determines chromatin domain boundaries, and how it contributes to disease is unclear (*4*). Expanded CAG/CTG repeat loci are ideal to address these questions because they are disease-associated loci with changes in local chromatin structure, transcriptional output, and genetic instability (*5*).

CAG/CTG repeats underlie 13 different neurological and neuromuscular disorders including myotonic dystrophy type 1 (DM1) and Huntington's disease (HD). They are part of a larger group of diseases caused by the expansion of short tandem repeats (STRs) (*6*). Disease-associated STRs (daSTRs) are genetically unstable, especially once they surpass a critical threshold of about 35 to 50 repeat units. Their expansion is also associated with extensive chromatin remodeling of the expanded loci (*5*, *7*). Two examples of diseases that are accompanied by extensive changes in chromatin marks are fragile X syndrome (FXS), caused by the expansion of CGG repeats in the *FMR1* gene located on the X chromosome (*8*–*10*), and Friedreich's ataxia (FRDA), caused by a homozygous GAA repeat expansion in

the first intron of the *FXN* gene (*11*). In the case of FXS, expansions beyond 200 CGGs are associated with promoter silencing in cis. This locus accumulates high levels of heterochromatic marks including CpG methylation, H4K20me3, H3K9me2/3, and H3K27me3 while losing euchromatin-associated marks such as H3 and H4 acetylation as well as H3K4me2 (*12*–*16*). Similar chromatin remodeling occurs at the *FXN* locus in FRDA patient tissues where expanded *FXN* alleles display increased DNA methylation, H3K9me2/3, H3K27me3, HP1 recruitment, and a loss of CCCTC-binding factor (CTCF) binding at sites flanking the expanded GAA repeats (*17*–*21*).

The shift from a euchromatic to a heterochromatic state upon daSTR expansion has led to the hypothesis that there is a concurrent change in the higher-order chromatin conformation of the surrounding genomic region. 3C-based experiments revealed alterations in 3D chromatin interactions surrounding expanded GAA and CGG repeats in patient-derived lymphoblastoid cell lines (LCLs) (*13*, *17*, *22*). In FRDA, a 3C anchor in exon 1 of *FXN* showed significantly higher interactions with genomic sites up to 39 kb upstream and 45 kb downstream of the GAA repeats in FRDA patient-derived LCLs compared to unaffected cells (*17*). In FXS patient postmortem brain tissue, fibroblasts, and LCLs, CGG repeat expansions were associated with the disruption of a topologically associating domain (TAD) boundary near the *FMR1* gene, decreased interactions near the repeat tract, and increased chromatin interactions within the upstream chromatin domain (*13*, *22*). Furthermore, daSTRs were found predominantly at TAD and sub-TAD boundaries enriched in CpG islands, suggesting more generally that daSTR expansions may disrupt TADs (*22*). Thus, it was speculated that altered chromatin conformation near expanded daSTR loci might contribute to repeat instability, transcriptional misregulation in cis, and ultimately to disease progression (*13*, *17*, *22*).

Two critical unknowns in this model are whether changes in higher-order chromatin structure are confined to CGG and GAA repeats or if this is general to daSTRs and whether changes in chromatin structure cause alterations in gene expression and repeat instability. The latter question is especially appealing in the context of the expanded CTG repeats in the 3′ untranslated region of the *DMPK* gene in DM1. Similar to *FMR1* and *FXN*, this locus undergoes

[1]Center for Integrative Genomics, Faculty of Biology and Medicine, University of Lausanne, 1015 Lausanne, Switzerland. [2]School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland. [3]Vital-IT Group, Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland. [4]Department of Genetic Medicine and Development, University of Geneva Medical School, 1211 Geneva, Switzerland. [5]Department of Computational Biology, Faculty of Biology and Medicine, University of Lausanne, 1015 Lausanne, Switzerland. [6]UK Dementia Research Institute at Cardiff University at Cardiff University, Hadyn Ellis Building, Maindy Road, CF24 4HQ Cardiff, UK. [7]Department of Molecular Mechanisms of Disease, University of Zurich, 8057 Zurich, Switzerland. [8]Institute for Genetics and Genomics in Geneva (iGE3), University of Geneva, 1211 Geneva, Switzerland.
*Corresponding author. Email: dionv@cardiff.ac.uk

heterochromatinization upon repeat expansion. The changes in local chromatin marks observed at the *DMPK* locus include the loss of CTCF binding, loss of a DNase I hypersensitive site, an increase in DNA and H3K9 methylation, as well as a loss of histone acetylation around the repeat tract (*23–27*). Moreover, the ectopic introduction of an expanded CAG repeat locus in budding yeast was sufficient to relocate the locus to the nuclear periphery in S phase cells, leading to changes in repeat size (*28*). Expanded CAG/CTG repeats cause most daSTR disorders. Together, these observations prompted us to test the hypothesis that chromatin conformation is altered at expanded CAG/CTG repeat loci. We used 4C sequencing (4C-seq) to determine the 3D chromatin interactions established around the *FMR1*, *HTT*, and *DMPK* loci in unaffected and FXS, HD, and DM1 patient LCLs. We also analyzed the chromatin interactions of an ectopic CAG repeat expansion in human embryonic kidney (HEK) 293–derived cells. We confirmed the alteration of chromatin interactions in a FXS patient cell line with ~935 CGG repeats. However, we found no evidence of changes in chromatin interactions caused by expanded CAG/CTG repeats in HD or DM1 patient cells. This suggests that 3D conformational changes are unlikely to underlie the alterations in transcriptional output or genetic instability of disease-associated expanded CAG/CTG repeat loci. This was consistent in different cell types, genetic backgrounds, and in the presence of specific heterochromatic marks. Therefore, we conclude that changes in higher-order chromatin conformation are unlikely to contribute to the pathogenesis of expanded CAG/CTG repeat disorders.

RESULTS
Chromatin conformation changes upon CGG repeat expansion at the *FMR1* locus
To determine the chromatin conformation of expanded daSTRs, we used 4C-seq (*29*) to maximize the resolution of chromatin interactions at these loci. This method allows a high sensitivity to small changes in conformation that may be missed by other 3C-based methods. To test our approach, we sought to reproduce the observation of changes in chromatin conformation at the *FMR1* locus upon CGG repeat expansion with 4C-seq (*13*, *22*). We used two LCLs that were used in a previous study (*22*), one derived from a patient with FXS with ~935 CGG repeats (GM09237) and the other derived from their unaffected male sibling (GM09236), here referred to as FXS and UN-A, respectively (Fig. 1A and table S1).

To determine the chromatin interactions established at the *FMR1* locus, we used a 4C viewpoint located 1 kb upstream of the CGG repeats and another at 195 kb upstream (table S2). We also determined the chromatin conformation of an unrelated locus, *ACTA1* on chromosome 1, to control for a potential effect of FXS on genome-wide chromatin conformation. We obtained three replicates of each 4C viewpoint and found that replicates from the same cell line showed good correlation in 4C fragments with at least 20 mapped reads (fig. S1, A to C). We identified interaction peaks and broader high-interacting regions that contacted the 4C viewpoints at frequencies higher than expected given their linear distance away from the viewpoints. These significant interactions were determined using two 4C-seq data analysis packages: FourCSeq (*30*) and 4C-ker (*31*). We then compared the chromatin interaction profiles between the UN-A and FXS cell lines to identify regions of significant differential interaction frequency (called differential interactions; see Materials

and Methods) (Fig. 1, B and C). As expected from 5C interaction maps (*22*), we found significant differences in chromatin interactions near the CGG repeats of *FMR1*, where the FXS patient cells showed decreased interactions encompassing the expanded repeats and the upstream regions of *FMR1* (Fig. 1B). By contrast, we did not find significant changes in chromatin interactions with a viewpoint located 195 kb upstream of *FMR1*, located within the upstream TAD (fig. S2). In addition, at the *ACTA1* locus, we found that the interaction profiles were similar between the two cell lines within a 2-Mb region around the viewpoint (Fig. 1C). This suggests that CGG repeat expansion alters chromatin interactions only in the vicinity of the *FMR1* locus. Thus, we confirmed previous observations made with 5C and we concluded that changes in chromatin conformation caused by daSTRs could readily be detected with 4C-seq.

Chromatin conformation is stable upon CAG repeat expansion at the *HTT* locus
To assess whether chromatin interactions change upon CAG repeat expansion at the *HTT* locus, we used three HD patient–derived LCLs: GM02164 (44 and 56 CAGs), GM03620 (18 and 70 CAGs), and GM14044 (19 and 750 CAGs), referred to as HD-A, HD-B, and HD-C, respectively. We compared them to two cell lines from unaffected individuals (GM04604 and GM02180, UN-B and UN-C, respectively). Their family relationships and their repeat sizes are shown in Fig. 2A and table S1.

To determine the chromatin conformation of the *HTT* locus, we used two 4C viewpoints within the *HTT* gene body—1 kb (HTT_d1) and 85 kb (HTT_d85) downstream of the CAG repeats in unaffected and HD patient cells (table S2). We also assessed the chromatin conformation of the unrelated *ACTA1* gene. Replicates from the same cell lines showed good correlation in fragments with more than 20 mapped reads (fig. S1, D to F). Compared to the UN-B and UN-C cells, we found that the chromatin interaction profiles were similar within a 2-Mb region around the *HTT* viewpoints in all three HD patient LCLs (Fig. 2B and fig. S4). The *ACTA1* viewpoint also produced indistinguishable interaction profiles between unaffected and HD patient cells (Fig. 2C). We identified a few small regions displaying differential interactions, but they were mainly outside regions of significant interactions (Fig. 2, B and C). In addition, most regions of differential interactions were not exclusive to HD patient cells, as they were also found in comparisons between the two unaffected cell lines (Fig. 2, B and C). This suggests that the minor changes in chromatin interaction frequencies are due to factors other than the presence of expanded CAG repeats in *HTT*, for example, a difference in genetic background. Together, our results show that expanded CAG repeats at the *HTT* locus do not cause significant alterations of its chromatin conformation.

To determine the relationship, if any, of chromatin conformation at the *HTT* locus with gene expression in cis, we performed RNA sequencing (RNA-seq) experiments with the same unaffected and HD patient cells. Differential gene expression analysis between the HD and unaffected LCLs identified 1183 significantly up-regulated and 1307 significantly down-regulated genes ($\log_2$ fold change > 0.5, adjusted $P < 0.05$; fig. S5A and table S3). Within a 2-Mb region centered around *HTT*, we identified 4 genes of 25 that showed significant differential expression between the HD patient and the unaffected cell lines (fig. S5A). This proportion was not statistically different from the proportion of differentially expressed genes in a 2-Mb region around *ACTA1* (Fisher's exact test, $P = 0.12$). Given that the
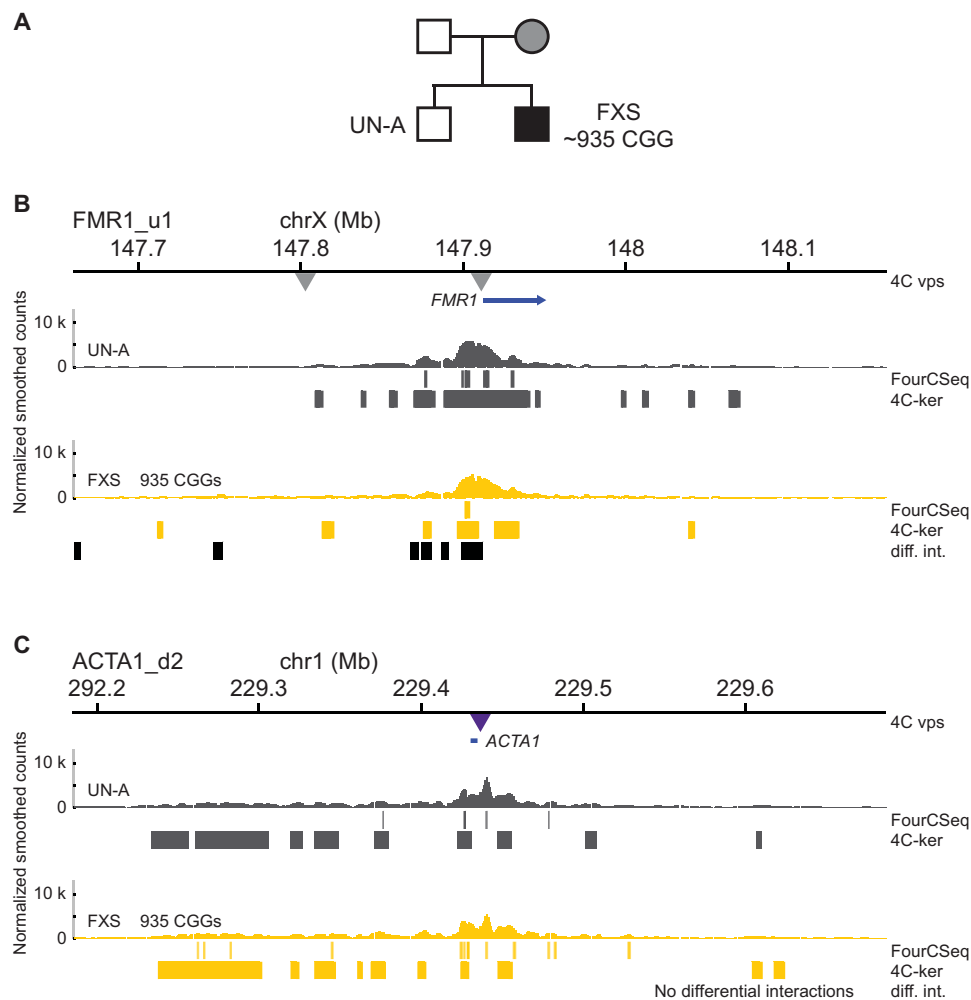
Fig. 1. Chromatin interactions of the *FMR1* locus in unaffected and FXS patient cells. (**A**) Pedigree of the unaffected and FXS patient cell lines used. (**B**) 4C-seq chromatin interaction profiles (average of triplicate smoothed and normalized counts) from the FMR1_u1 viewpoint (1 kb upstream of the CGG repeats of *FMR1*, gray central triangle) in one unaffected (UN-A) and one FXS patient cell lines (FXS). The top blue arrow represents the *FMR1* gene and the left-side triangle represents the location of the FMR1_u195 4C viewpoint. The interaction profiles for the FMR1_u195 viewpoint (195 kb upstream of the CGG repeats of *FMR1*) can be found in fig. S2. (**C**) 4C-seq chromatin interaction profiles (average of triplicate smoothed and normalized counts) from the *ACTA1* viewpoint (central purple triangle). The top blue bar represents the *ACTA1* gene. For (B) and (C), high-interacting regions were called using 4C-ker and significant interactions were called using FourCSeq. Regions of differential interactions compared to UN-A are marked with black bars below each 4C-seq track and labeled as "diff. int.". 4C vps, viewpoints used for 4C-seq.

chromatin interactions did not significantly change between the HD and unaffected cell lines in the *HTT* region, our data suggest that changes in gene expression in cis occur independently of the chromatin conformation of expanded *HTT* alleles in HD patient cells.

## Chromatin conformation is stable upon CTG repeat expansion at the *DMPK* locus

Repeat expansion at the *HTT* locus is not known to be associated with changes in local histone modifications and chromatin accessibility, which might cause changes in chromatin conformation. To determine whether an expanded CTG repeat locus with local heterochromatic chromatin marks is associated with altered 3D chromatin conformation, we used the unaffected cell lines UN-B and UN-C and two DM1 patient–derived LCLs (GM06077 and GM04648, referred to as DM1-A and DM1-B, respectively). The DM1-A cell line harbored one expanded *DMPK* allele with 1700 CTGs and the DM1-B cell line, 1000 CTGs (Fig. 3A and table S1). Using bisulfite

sequencing, we found that DM1-A cells harbored increased CpG methylation levels at two CTCF binding sites flanking the CTG repeats (fig. S6, A and B). To assess the CTCF occupancy levels at the two flanking sites, we performed CTCF chromatin immunoprecipitation (ChIP) followed by quantitative PCR (qPCR) and found that CTCF binding was reduced at both sites in the DM1-A cell line (fig. S6C). In DM1-B cells, we observed normal methylation levels at both CTCF binding sites (fig. S6B) and slightly reduced CTCF occupancy in the downstream site (fig. S6C). Thus, DM1-A cells displayed molecular signatures of congenital DM1 (*32*), whereas DM1-B cells had characteristics of adult-onset DM1 (table S1).

To determine whether expanded CTG repeats affect chromatin conformation at the *DMPK* locus, we performed 4C-seq with the unaffected and DM1 cell lines using four different 4C viewpoints at distinct distances away from the CTG repeats of *DMPK* (Fig. 3B, fig. S7, and table S2). Replicates from the same cell lines also showed good correlation for 4C fragments with at least 20 reads (fig. S1, G to K).
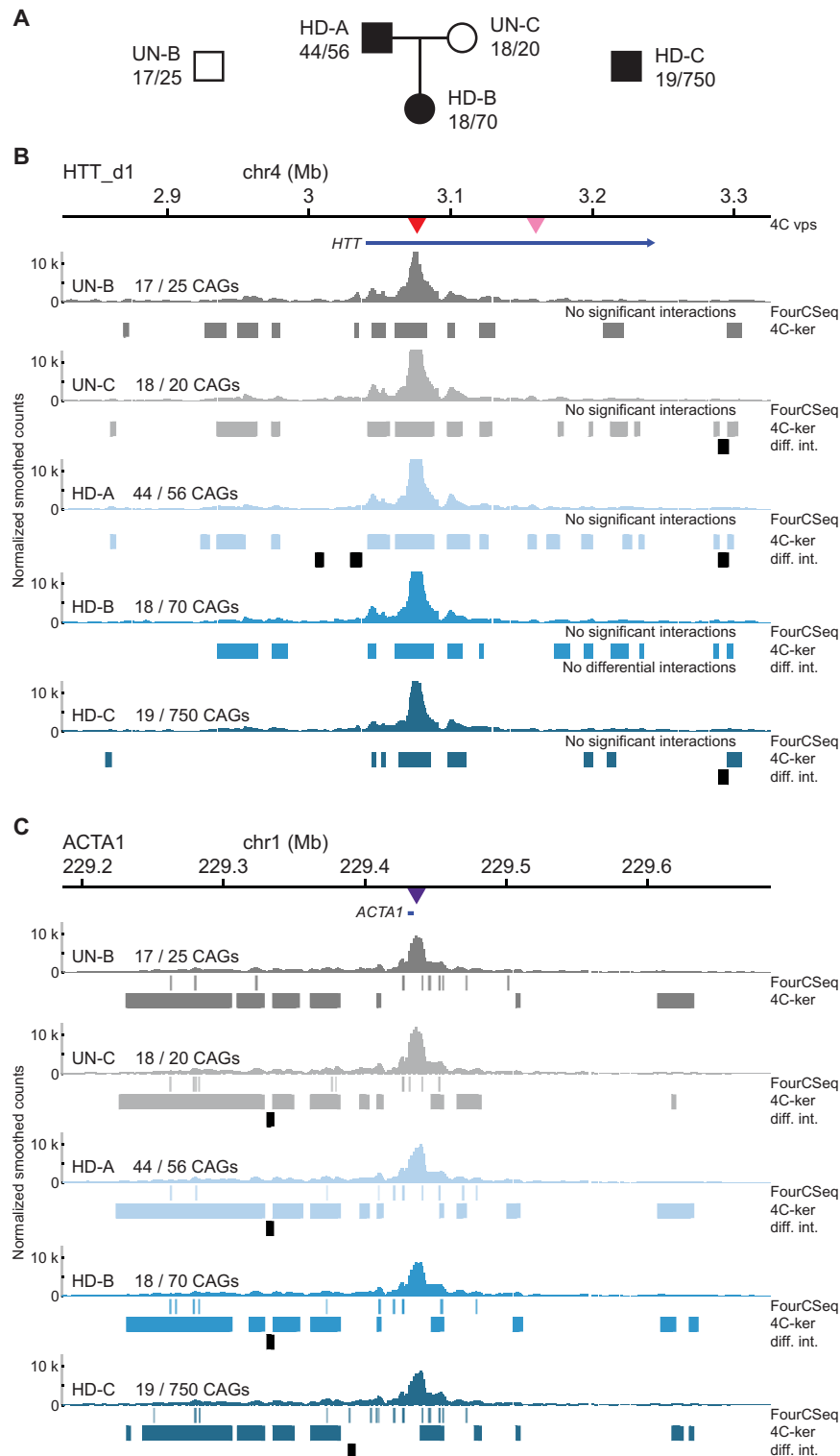
Fig. 2. Chromatin interactions of the *HTT* locus in unaffected and HD patient cells. (**A**) Pedigree of the unaffected and HD patient cell lines used. (**B**) 4C-seq chromatin interaction profiles (average of triplicate smoothed and normalized counts) from the HTT_d1 viewpoint (1 kb downstream of the CAG repeats of *HTT*, red central triangle) in two unaffected (UN-B and UN-C) and three HD patient cell lines (HD-A, HD-B, and HD-C). The top blue arrow represents the *HTT* gene and the triangles represent the location of both *HTT* viewpoints. The interaction profiles for the HTT_d85 viewpoint (85 kb downstream of the CAG repeats) can be found in fig. S4. (**C**) 4C-seq chromatin interaction profiles (average of triplicate smoothed and normalized counts) from the *ACTA1* viewpoint (central purple triangle). The top blue bar represents the *ACTA1* gene. For (B) and (C), high-interacting regions were called using 4C-ker and significant interactions were called using FourCSeq. Regions of differential interactions compared to UN-B are marked with black bars below each 4C-seq track and labeled as "diff. int.".
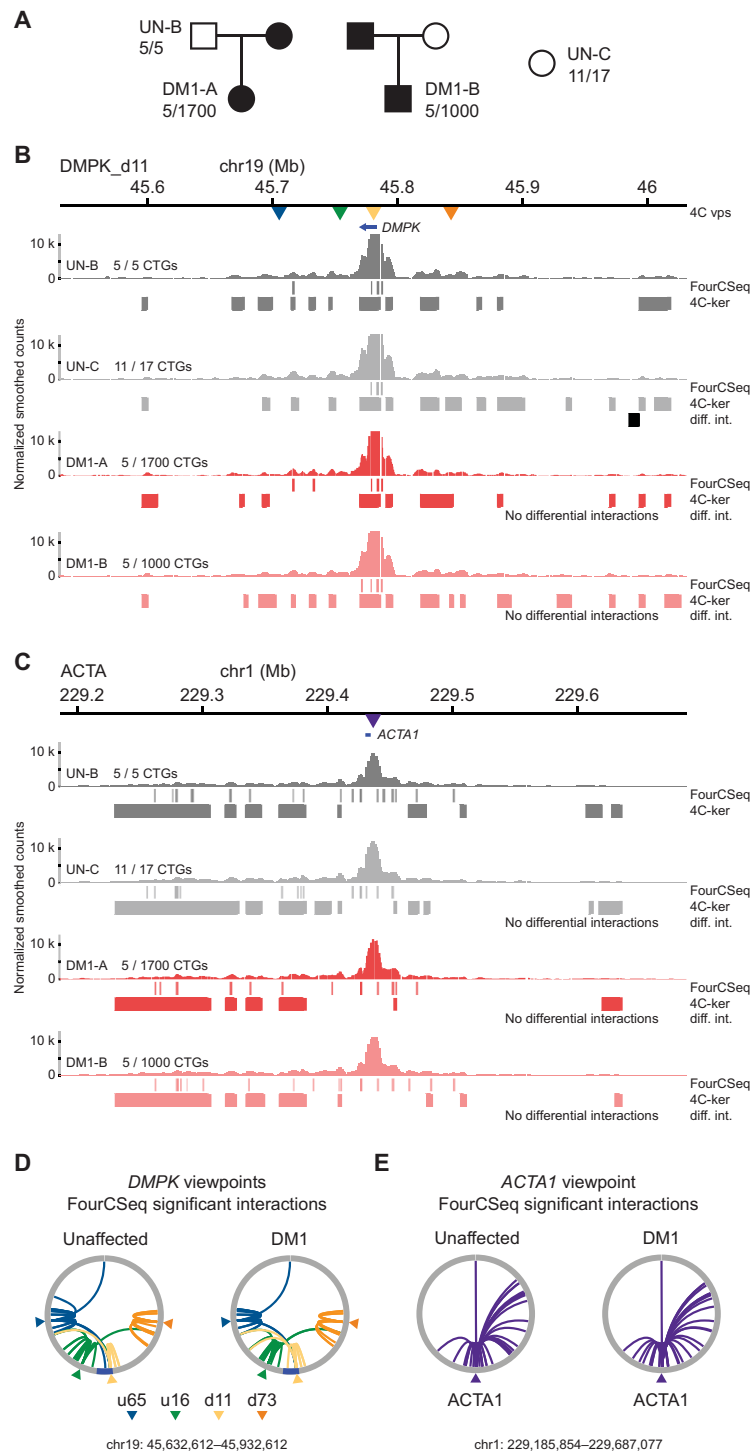
Fig. 3. Chromatin interactions of the *DMPK* locus in unaffected and DM1 patient cells. (**A**) Pedigree of the unaffected and DM1 patient cell lines used. (**B**) 4C-seq chromatin interaction profiles (average of triplicate smoothed and normalized counts) from the DMPK_d11 viewpoint (11 kb downstream of the CTG repeats of *DMPK*, yellow triangle) in two unaffected (UN-B and UN-C) and two DM1 patient cell lines (DM1-A and DM1-B). The top blue arrow represents the *DMPK* gene and the triangles represent the location of the four *DMPK* viewpoints. The profiles for the three other viewpoints can be found in fig. S7. (**C**) 4C-seq chromatin interaction profiles (average of triplicate smoothed and normalized counts) from the *ACTA1* viewpoint (central purple triangle) in two unaffected and two DM1 patient cell lines. The top blue bar represents the *ACTA1* gene. For (B) and (C), high-interacting regions were called using 4C-ker and significant interactions were called using FourCSeq. Regions of differential interactions compared to UN-B are marked with black bars below each 4C-seq track and labeled as "diff. int.". (**D**) Circos plot of the significant interactions called with FourCSeq (nominal $P < 0.05$) from four different viewpoints surrounding the CTG repeats of *DMPK* in the unaffected and DM1 cell lines (left and right, respectively) (DMPK_u65 in blue, DMPK_u16 in green, DMPK_d11 in yellow, and DMPK_d73 in orange; 65 kb upstream, 16 kb upstream, 11 kb downstream, and 73 kb downstream, respectively). (**E**) Circos plot of the significant interactions called with FourCSeq (nominal $P < 0.05$) from the control *ACTA1* viewpoint in unaffected and DM1 cell lines (left and right, respectively).

Similar to the HD samples, we observed notably similar chromatin interaction profiles between the unaffected and DM1 samples for all four viewpoints in the *DMPK* region (Fig. 3B and fig. S7). As expected, the *ACTA1* viewpoint also showed interaction profiles indistinguishable between unaffected and DM1 cell lines (Fig. 3C). As with the *HTT* viewpoints, none of the *DMPK* or *ACTA1* viewpoints in DM1 patient cells had significant interactions that were also called as regions of differential interaction (Fig. 3, B and C). Furthermore, we found that the significant interactions identified for all four *DMPK* viewpoints and the *ACTA1* viewpoint in unaffected and DM1 patient cells were largely the same (Fig. 3, D and E). Thus, we found no evidence of large-scale changes in chromatin conformation at the *DMPK* locus driven by CTG expansions.

We confirmed by RNA-seq that transcriptional misregulation in DM1 patient cells occurs genome-wide as well as within a 2-Mb region centered around *DMPK* (fig. S5B and table S4). Similar to our observation in HD individuals, changes in gene expression in cis occurred independently of the 3D conformation of the *DMPK* locus in DM1 patient cells.

To determine the extent of CTCF occupancy alterations in the DM1 patient cells, we performed ChIP sequencing (ChIP-seq) (fig. S8). At the genome-wide level, comparing DM1-A to UN-B cells, we found 1951 sites where CTCF binding changed significantly [false discovery rate (FDR) < 5%]. DM1-B cells harbored 833 sites with significantly altered CTCF occupancy genome-wide compared to UN-B cells (FDR < 5%). However, in a 2-Mb region centered around *DMPK*, only 4 of 131 (3%) CTCF binding peaks showed significant differential occupancy in DM1-A patient cells compared to the UN-B cell line (fig. S8). There were no binding sites with significant differences in CTCF occupancy out of 135 peaks in that same region when comparing the UN-B and DM1-B cell lines (fig. S8). Therefore, we concluded that there were no large-scale changes in CTCF occupancy in the DM1 patient LCLs either genome-wide or within a 2-Mb region around the expanded CTG repeats, which is consistent with the absence of changes in chromatin conformation observed by 4C-seq in these cells.

Lack of allelic bias in chromatin interactions at expanded CAG/CTG repeats

DM1 and HD are both dominantly inherited disorders wherein affected individuals are heterozygous for the expanded allele. Thus, one potential caveat in our data was that the presence of a normal-length allele could mask changes in the 3D chromatin interactions made by the expanded allele. To evaluate this possibility, we took advantage of the presence of at least one parental cell line for the DM1-A and HD-B individuals in our dataset (Figs. 2A and 3A) and identified biallelic single-nucleotide polymorphisms (SNPs) in the contacted 4C fragments from these samples. We selected a subset of the SNPs that could be assigned unambiguously to either the expanded or the normal allele within 1 Mb of the 4C viewpoints (see Materials and Methods). We reasoned that if chromatin contacts were established without a systematic bias for either the expanded or normal allele in DM1 and HD patient cells, then the proportion of 4C fragments in which the expanded chromosome had more reads than the normal-length one would be close to 50%. Therefore, we analyzed the sequencing coverage of the 4C-seq data at the biallelic SNP positions and found that the viewpoints did not establish preferential contacts with a single chromosome in either the DM1-A or HD-B patient cell lines (Table 1). These results are consistent with the conclusion that chromatin interactions at both disease loci do not show allelic bias. Together, these results corroborate the conclusion that expanded CAG/CTG repeats do not significantly alter the chromatin interactions at two expanded CAG/CTG repeat loci.

Chromatin conformation is stable at an ectopic CAG repeat locus

It remained possible that differences in genetic background in the LCLs that we analyzed could have had a confounding effect on the chromatin interactions made at expanded CAG/CTG repeat loci. To test this, we compared the chromatin interactions of a hemizygous ectopic locus with either 15 [GFP(CAG)$_{15}$] or 270 CAGs [GFP(CAG)$_{270}$] in isogenic cell lines. We obtained two clonal populations of HEK293 T-Rex Flp-In cells that contain a single, stably integrated construct containing CAG repeats within the intron of a green fluorescent protein (GFP) mini-gene controlled by a doxycycline-inducible promoter (*33*). Using targeted locus amplification (*34*), we mapped its insertion site to the p-arm of chromosome 12, 1.2 Mb from the telomere (Fig. 4A). We performed 4C-seq in both cell lines with a viewpoint located 1 kb upstream of the CAG repeats. The chromatin interaction profiles of this ectopic CAG repeat locus were similar between the 15 and 270 CAG repeat cells, with few regions of differential interactions overlapping with high-interacting regions identified with 4C-ker (Fig. 4B). Thus, we concluded that expanded CAG repeats cause few changes to the chromatin conformation of an ectopic locus in isogenic cells.

Table 1. Allele-specific interactions in 4C-seq data.

Cell line	Disease	Genomic region*	Total no. of samples	No. of samples with more normal chromosome reads	No. of samples with more expanded chromosome reads	P†
DM1-A	DM1	*ACTA1*	5	1	4	0.38
		DMPK	41‡	26	15	0.12
HD-B	HD	*ACTA1*	12	7	5	0.77
		HTT	43	20	23	0.76

*One-megabase region around the 3′ end of *DMPK*, 5′ end of *HTT*, or the *ACTA1* 4C-seq viewpoint. †*P* value of an exact binomial test. ‡One sample had the same number of reads in both alleles; thus, it was not included.
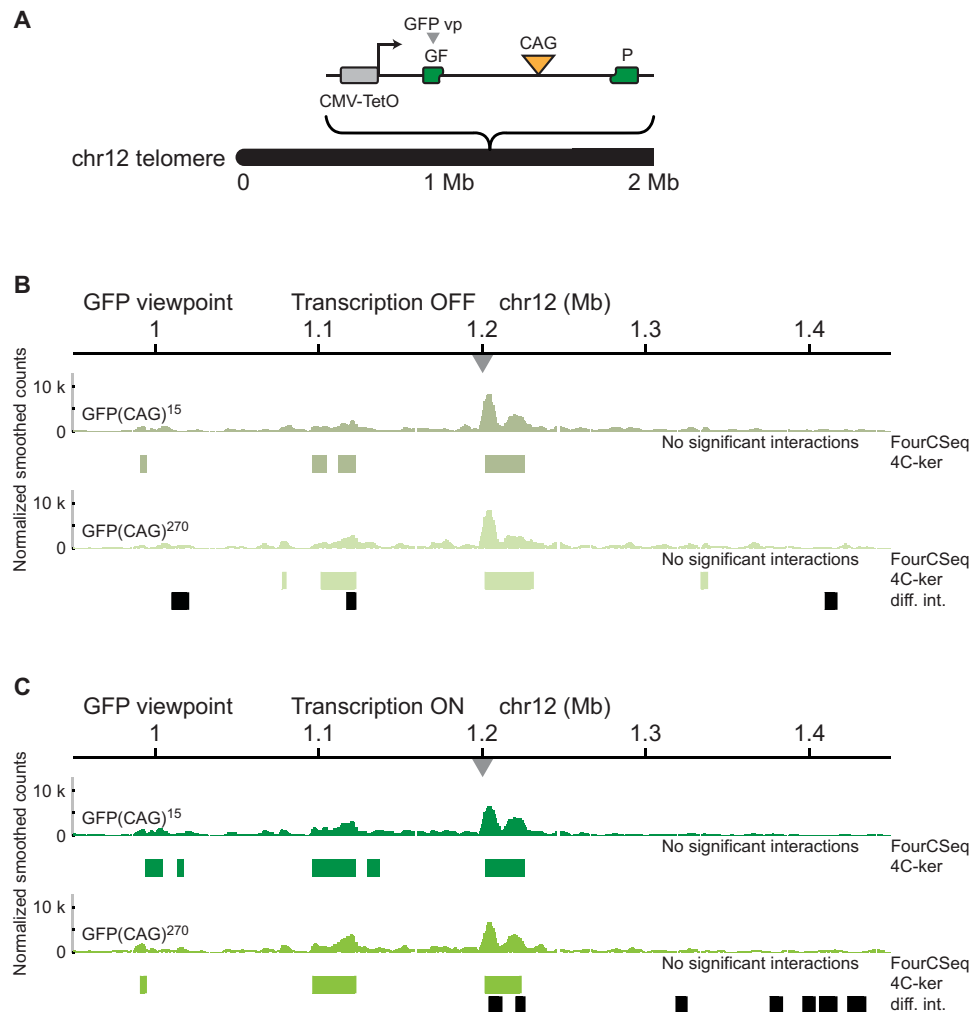
Fig. 4. Chromatin interactions of an expanded CAG repeat locus in isogenic cells. (**A**) Diagram of the integration site of the CAG ectopic locus in GFP(CAG)$_{15}$ and GFP(CAG)$_{270}$ cells. (**B**) 4C-seq chromatin interaction profiles (average of triplicate smoothed and normalized counts) from the GFP viewpoint in GFP(CAG)$_{15}$ (top) and GFP(CAG)$_{270}$ (bottom) in cells without active transcription of the ectopic CAG locus. Regions of differential interactions compared to GFP(CAG)$_{15}$ are marked with black bars below the GFP(CAG)$_{270}$ 4C-seq track. (**C**) 4C-seq chromatin interactions from the GFP viewpoint in GFP(CAG)$_{15}$ (top) and GFP(CAG)$_{270}$ (bottom) in cells treated with doxycycline for 5 days to induce transcription of the ectopic CAG locus. Regions of differential interactions compared to GFP(CAG)$_{15}$ are marked with black bars below the GFP(CAG)$_{270}$ 4C-seq track.

Some studies suggest that transcription can help define chromatin domain boundaries (*35*). To determine whether transcription through expanded CAG repeats could lead to changes in chromosome conformation, we induced transcription at this locus by culturing the GFP(CAG)$_{15}$ and GFP(CAG)$_{270}$ cells with doxycycline for 5 days and subsequently performed 4C-seq using the same GFP viewpoint. We observed that 2 of 17 regions of differential interactions overlapped fully or in part with high-interacting regions over a 2-Mb region (Fig. 4C). By contrast, we identified no significant interactions in this region using FourCSeq. Together, these results suggest that an ectopic expanded CAG repeat tract is not enough to significantly alter the chromosome conformation of this locus, regardless of its transcriptional status.

DISCUSSION
Here, we showed that chromatin interactions remain stable at expanded CAG/CTG repeat loci. This was true for two disease loci (*HTT*

and *DMPK*) and one ectopic locus in two different cell types. Our findings are supported by allele-specific analysis of 4C-seq chromatin interactions. Furthermore, increased CpG methylation and CTCF binding alterations at four sites in a 2-Mb region around *DMPK* did not affect the chromosome conformation of this locus in the DM1-A patient cells. This is especially relevant because CTCF is a key architectural protein involved in the demarcation of TAD and sub-TAD boundaries (*36*). In addition, when we inserted a hemizygous transgene with CAG repeats, we found that expanded CAG repeats did not induce a significant reorganization of the chromatin contacts established at this ectopic repeat locus. These results show that an expanded CAG/CTG repeat tract is not sufficient to change chromatin conformation in cis.

One possible explanation for why we observed strikingly similar chromatin interactions in the expanded CAG/CTG repeat loci tested here is that changes in chromatin interactions caused by expanded daSTRs only occur in affected cell types, such as cardiomyocytes or medium spiny neurons, but not in LCLs. Such a scenario would

require cell type–specific factors to mediate these changes. Here, we tested the transcriptional status, DNA methylation levels, and CTCF binding in DM1 patient LCLs. We found that these cells displayed transcriptional and epigenetic signatures similar to those found in heart, liver, cortex, and peripheral blood samples of patients with DM1 (*23*, *32*). Thus, cell type–specific differences in chromatin conformation are possible but unlikely given the similarity in local chromatin modifications between the tissues.

Our findings are in contrast to the effect of expanded CGG and GAA repeats on chromosome conformation (*13*, *17*, *22*). In FXS patient–derived LCLs, fibroblasts, and brain tissue, expanded CGG repeats in *FMR1* were associated with heterochromatic characteristics, decreased CTCF binding, and a disruption of a TAD boundary near the expanded repeats (*22*). The DM1 patient cells used here, especially the congenital DM1 patient cell line (DM1-A), showed similar changes in chromatin modifications to FXS patient cells. And yet, these factors did not amount to an alteration of the 3D chromatin interactions at the expanded *DMPK* locus. There are notable dissimilarities between FXS/FRDA and DM1/HD that could account for the difference in the effect of distinct daSTRs on chromosome conformation. Apart from the nucleotide composition of the repeat tract itself, the disease loci and their flanking sequences are different. It is possible that the extent of heterochromatin formation is different between distinct expanded repeat loci, thereby eliciting different effects on the surrounding 3D chromatin conformation. Similarly, how far CpG methylation spreads and how much CTCF occupancy is disrupted upon repeat expansion is not fully understood and may account for the differences between the daSTR loci examined so far. It seems likely that all these dissimilarities between the loci examined so far contribute to the changes, or lack thereof, in chromatin conformation upon expansion of daSTRs.

It was speculated that the changes in chromatin conformation caused by expanded daSTRs could lead to repeat instability (*22*). CGG repeat expansions beyond 200 units are associated with reduced instability, promoter silencing, and 3D chromatin conformation changes (*22*, *37*, *38*). By contrast, long GAA repeat tracts are more unstable (*6*, *39*), and yet, their chromosome conformation is also altered. We found that long CAG/CTG repeats, which are more unstable as they expand, are not associated with changes in 3D chromatin interactions. Thus, there does not appear to be a general functional link between chromatin conformation and repeat instability across distinct daSTR loci.

Similarly, we found that transcription around the *HTT* and *DMPK* regions in patients with HD and DM1 is unrelated to the chromatin conformation of the expanded repeat loci. Rather, it is more likely that the local transcriptional misregulation observed in patients is caused by changes in local chromatin marks. Our results are in line with recent evidence in *Drosophila* arguing that genome topology is not predictive of genome-wide transcriptional output (*40*). We did not find evidence that supports a common link between 3D genome organization and the transcriptional misregulation typical of expanded daSTRs. Overall, our data argue that changes in chromatin topology are unlikely to underpin the molecular pathology of expanded CAG/CTG repeat disorders.

The expanded CAG/CTG repeat loci in DM1 and HD provide an endogenous genomic substrate to study the mechanisms necessary for the establishment of 3D chromatin domains in daSTRs given the involvement of CTCF binding, CpG methylation, and other chromatin remodeling events at these loci. In particular, CTCF plays a central role in the establishment of chromatin loops and chromatin contact domains (*36*). Our data show that chromatin interactions remain stable near the *DMPK* region despite precise alterations in CTCF binding near a chromatin contact domain boundary. This is in line with several recent observations wherein interfering with CTCF binding at chromatin domain boundaries does not result in a reorganization of the underlying 3D topological chromatin structure. For example, neither the deletion nor the ectopic insertion of the *Firre* locus, a long noncoding RNA on the X chromosome surrounded by 15 CTCF binding sites, was enough to alter the TAD structure of the surrounding genomic region in mouse cells (*41*). Similarly, a TAD boundary in the *HoxD* gene cluster was highly resilient to genomic deletions encompassing multiple CTCF binding sites (*42*). Therefore, these results, together with ours, imply a model whereby the establishment and maintenance of chromatin conformation is dependent on the chromosomal context of a given genomic locus.

MATERIALS AND METHODS
Cell lines
All LCLs were obtained from the Coriell Institute for Medical Research Cell Repository. They were grown at 37°C and 5% CO_2 in RPMI 1640 medium supplemented with 15% fetal bovine serum, 2 mM L-glutamine, and 1% penicillin-streptomycin. Cells were counted and passaged every 3 to 4 days depending on cell density. The GFP(CAG)$_{270}$ and GFP(CAG)$_{15}$ lines were previously characterized (*33*). They were maintained at 37°C and 5% CO_2 in Dulbecco's minimum Eagle's medium with GlutaMAX, 1% penicillin-streptomycin, blasticidine (15 µg ml^{-1}), and hygromycin (150 µg ml^{-1}). Transcription of the GFP mini-gene was activated by culturing GFP(CAG)$_n$ cells with doxycycline at a final concentration of 2 µg ml^{-1} for 5 days.

Repeat length determination and small-pool PCR
Genomic DNA was isolated from each LCL using the NucleoSpin Tissue Kit (Macherey-Nagel). PCR products with the CAG repeats from *HTT* were produced with primers oVIN-1333 and oVIN-1334 (table S5). PCR products containing the CTG repeats from *DMPK* were amplified with primers oVIN-1252 and oVIN-1251 (table S5). For normal-length alleles, several PCRs were set up with MangoTaq (Bioline), and the products were gel-extracted and Sanger-sequenced with the same primers used for the amplification. For expanded alleles, small-pool PCRs were performed on the basis of a previously described protocol (*43*). Briefly, the same primers were used for the amplification of expanded *DMPK* and *HTT* alleles with 1 ng of genomic DNA per PCR. The products were run on an agarose gel and transferred to a nylon membrane. An oligo made up of 10 CAGs was used to obtain a radioactive probe used for the visualization of the expanded alleles. The number of repeats reported here is an estimation of the modal number of repeats.

Bisulfite sequencing
Bisulfite sequencing was performed according to a previously described method (*32*). Bisulfite conversion was performed with the EZ DNA Methylation Kit (Zymo Research) using the standard protocol. Two hundred nanograms of genomic DNA was used for bisulfite conversion at 50°C for 12 hours. Bisulfite-converted DNA was desulfonated, eluted, and immediately used for nested and heminested PCR amplification of the upstream and downstream CTCF binding sites, respectively, using previously described primers (*32*). Fifty nanograms

of bisulfite-converted DNA was used for the first PCR, and 3 μl of the products was used for the second PCR. The final amplicons were purified with the NucleoSpin PCR Clean-up Kit (Macherey-Nagel) and used for 2 × 250–base pair (bp) paired-end sequencing on Illumina MiSeq. The primers used for both rounds of PCR are found in table S5.

Bisulfite sequencing data analysis

Sequencing reads were preprocessed using Trim Galore (github.com/FelixKrueger/TrimGalore) with the following parameters: -q 20 --length 20 –paired. Reads were aligned using QuasR (44) to the GRCh38 human reference DNA sequences corresponding to the amplified PCR products. DNA methylation calls for each CpG were extracted using the qMeth() function in QuasR. DNA methylation frequencies were calculated as methylated CpG reads / total number of reads covering the respective CpG × 100.

CTCF ChIP-qPCR and ChIP-seq

ChIP was performed according to the Diagenode Auto iDeal ChIP-qPCR Kit (Diagenode) standard protocol. Samples of 6×10^6 cells were sonicated using the Diagenode Bioruptor Pico (Diagenode), with 10 cycles of 30 s "on" and 30 s "off." Correct DNA fragmentation was verified by agarose gel electrophoresis. Immunoprecipitation was performed with 4×10^6 cells using a CTCF antibody (Diagenode) and the Diagenode IP-Star Compact Automated System robot (Diagenode). Results were analyzed using the StepOnePlus qPCR by Applied Biosystems (Thermo Fisher Scientific) with Applied Biosystems SYBR Green PCR Master Mix (Thermo Fisher Scientific). The primer sequences used for qPCR are listed in table S5. CTCF ChIP-seq libraries were prepared in triplicate using the Illumina TruSeq ChIP Sample Preparation Kit. We performed 1 × 50-bp single-end sequencing with Illumina HiSeq 4000.

CTCF ChIP-seq data analysis

CTCF ChIP-seq reads were demultiplexed with the bcl2fastq2 Illumina software (v 2.20.0). Data analysis was performed according to the ENCODE Transcription Factor ChIP-seq processing pipeline (phase 3). Demultiplexed reads were aligned to the GRCh38 human reference genome (GCA_000001405.15) using bowtie2 (v 2.3.4.3) (45). Reads were filtered to remove unmapped reads, reads with no primary alignment, multimapped reads, and duplicate reads so that only the reads that mapped to the genome once were considered for downstream analyses. CTCF peaks were identified using the ChIP-seq processing pipeline (46). The irreproducible discovery rate (47) framework was used to ensure reproducibility between experimental replicates. Differentially enriched CTCF peaks were detected with DiffBind (48). We used DESeq2 (49) to identify peaks that are statistically differentially bound between sample groups ($P < 0.05$ and FDR < 5%). First, we computed count information for each of the peaks in the consensus set. Bedgraph files (representing a normalized coverage per base) were created to visualize the tracks.

4C sequencing

4C library preparation was performed on the basis of a previously described protocol (29). For each sample, 10^7 cells were cross-linked in 2% formaldehyde for 10 min at room temperature and quenched with glycine to a final concentration of 0.13 M. Cross-linked samples were rinsed once with phosphate-buffered saline and were either used immediately or flash-frozen in liquid nitrogen and stored at −80°C for later use. Cells were lysed for 15 min on ice in lysis buffer [50 mM

tris-HCl (pH 7.5), 150 mM NaCl, 5 mM EDTA, 0.5% NP-40, 1% Triton X-100, and 1× protease inhibitors]. The first digestion was performed with 200 U of Dpn II (New England Biolabs) and incubating for 4 hours at 37°C, then 200 U of Dpn II and overnight incubation at 37°C, and lastly 200 U of Dpn II and incubating 4 hours at 37°C. Dpn II digestion efficiency was assessed by agarose gel electrophoresis, and Dpn II was subsequently heat inactivated. The first ligation was performed at 16°C overnight with 50 U of T4 DNA ligase (Thermo Fisher Scientific) in 7 ml. Ligated samples were decross-linked with 30-μl Proteinase K (10 mg/ml) at 65°C overnight followed by ribonuclease A (RNAse A) treatment and phenol-chloroform purification. The second digestion was performed with 50 U of Bfa I (New England Biolabs) at 37°C overnight followed by Bfa I heat inactivation. The second ligation was performed at 16°C overnight with 100 U of T4 DNA ligase in 14 ml. The resulting 4C template samples were precipitated and purified with the QIAquick PCR purification Kit. 4C libraries were generated by amplifying 1-μg total of purified 4C template using the Expand Long Template PCR System (Roche) with 4C viewpoint primers with Illumina sequencing adapters that permitted multiplexing (in 50-μl PCRs), pooling reactions, and purifying the PCR products with AMPure XP beads to exclude products less than 130 bp. The 4C viewpoint primer sequences are listed in table S5. Single-end sequencing of pooled 4C libraries was performed on Illumina HiSeq 2500.

4C-seq data analysis

Demultiplexing, trimming, and mapping were performed using the BBCF HTSstation (50), according to (51). Reads were trimmed to keep the first 40 bp. The 4C fragments surrounding the viewpoints (±2.5 kb) were excluded from the rest of the analysis. The demultiplexed reads were mapped to the GRCh38 human reference genome using bowtie2 (v 2.2) (45). Fragment read counts were obtained using FourCSeq (v 1.18.0) (30). The number of mapped reads for each sample is found in table S2. For plotting the data, fragment counts were normalized (reads per million) and smoothed with a running mean (window size = 5 fragments). The smoothed and normalized fragment counts were averaged among replicates of the same 4C library samples and visualized with gFeatBrowser (www.gfeatbrowser.com). Significant chromatin interactions were identified with two 4C-seq data analysis packages: FourCSeq (v 1.18.0) (30) and 4C-ker (v 0.0.0.9000) (31). For the FourCSeq analysis, we defined significant interactions as fragments with a z score equal to or greater than 1.96 and an FDR of 0.1, using the following parameters: minCount = 20 and fitFun = "distFitMonotone" in the getZScores function; zScoreThresh = 1.96, fdrThresh = 0.1 in the addPeaks function. 4C-ker uses a Hidden Markov Model that accounts for differences in coverage near 4C viewpoints to determine three types of domains: high-interacting, low-interacting, and noninteracting domains. For each viewpoint, we used $k = 5$ in the nearBaitAnalysis function and plotted the high-interacting regions. The difference between significant interactions called with FourCSeq and 4C-ker is expected given that FourCSeq usually identifies "peaks" of significantly interacting regions whereas 4C-ker identifies regions (52). Differential interactions were identified with the differentialAnalysis function of 4C-ker, which is based on the DESeq2 (49) framework, using default parameters (including a P value threshold of 0.05). The UN-B cell line was used as the reference condition for all comparisons except for Fig. 1 and fig. S2 where the reference condition was the UN-A cell line.

4C SNP data analysis

We called SNPs from 4C-seq data from the DM1-A and HD-B patient–derived LCLs using GATK (v 3.7.0) (53) and samtools/bcftools (v 1.5 and v 1.4.1, respectively) (54). Biallelic SNPs located within a 1-Mb region of the 3′ end of *DMPK*, the 5′ end of *HTT*, and the *ACTA1* 4C viewpoint were selected. Among these, only SNPs that could unambiguously inform which parental allele they came from were retained for downstream analysis. This required a homozygous genotype in at least one of the parental cell lines. We validated a subset of these SNPs by isolating genomic DNA from the UN-B, UN-C, HD-A, HD-B, and DM1-A cell lines and using it for PCR amplification of the genomic region encompassing the variants in the parental and offspring LCLs followed by library preparation and 2×250-bp paired-end sequencing was performed with Illumina MiSeq. We then analyzed the sequencing coverage at the confirmed biallelic SNP positions from the 4C-seq data from DM1-A and HD-B patient cell lines. For samples with at least 10 reads per SNP position, we counted the number of times the expanded allele had more mapped reads than the normal allele. We applied an exact binomial test to statistically assess whether the proportion of cases where the expanded allele had more reads than the normal allele was significantly different to 0.5, which represented the null hypothesis of no allelic bias.

RNA isolation and sequencing

For each sample, RNA from 5 million cells was extracted using the NucleoSpin RNA Kit (Macherey-Nagel), and its concentration was measured by NanoDrop (NanoPhotometer NP80, Implen). RNA quality was assessed on a Fragment Analyzer (Agilent Technologies), and all of the RNA samples had an RNA quality number between 8.7 and 10. RNA-seq libraries were prepared using 500 ng of total RNA and the Illumina TruSeq Stranded Total RNA Library Prep Kit with Ribo-Zero Gold (Illumina) according to the manufacturer's protocol. Multiplexed samples were pooled in equimolar amounts, and 2×100-bp paired-end sequencing was performed on Illumina HiSeq 2500.

RNA-seq data analysis

RNA-seq reads were demultiplexed using the bcl2fastq Illumina software (v 2.20.0). Purity-filtered reads were trimmed with Cutadapt (v 1.3) (55) and filtered for low complexity with seq_crumbs (v 0.1.8). Reads were aligned against the GRCh38 human transcriptome (GRCh38.82) using STAR (v 2.4.2a) (56). Genes differentially expressed in HD and DM1 LCLs compared to unaffected LCLs were determined using DESeq2 (v 1.20.0) (49) while accounting for batch effects including cell passage number, collection date, and sequencing runs. *P* values were adjusted using the Benjamini-Hochberg multiple testing correction. Significantly differentially expressed genes were defined as those with adjusted *P* values less than 0.05 and $\log_2$ fold change greater than 0.5. Functional gene ontology enrichment analyses of significantly differentially expressed genes in HD and DM1 LCLs were performed using the DAVID tool (v 6.8) (57, 58), and all of the expressed genes in HD and DM1, respectively, were used as the background.

Targeted locus amplification

Targeted locus amplification (TLA) was performed on the basis of a previously described protocol (34). For each sample, 10^7 cells were cross-linked in 2% formaldehyde for 10 min at room temperature and quenched with glycine to a final concentration of 275 mM. Cell were lysed for 5 min at room temperature in lysis buffer [50 mM tris-HCl (pH 7.5), 150 mM NaCl, 5 mM EDTA, 0.5% NP-40, 1% Triton X-100, and 1× protease inhibitors). Cross-linked samples were digested at 37°C overnight with 400 U of Nla III (New England Biolabs) followed by Nla III heat inactivation. Samples were ligated at room temperature for 2 hours with 20 U of T4 DNA ligase (Thermo Fisher Scientific) in 500 μl. Ligated samples were decross-linked with 5-μl Proteinase K (10 mg/ml) at 65°C overnight, followed by RNAse A treatment and phenol-chloroform purification. Samples then digested overnight at 37°C with 50 U of Nsp I (New England Biolabs) followed by Nsp I inactivation. The second ligation was performed overnight at 16°C with 100 U of T4 DNA ligase in 14 ml. The resulting TLA circularized templates were purified with the QIAquick PCR Purification Kit. TLA libraries were generated by amplifying 800 ng of purified TLA template with TLA viewpoint primers (table S5). Paired-end sequencing (2×150 bp) of pooled TLA libraries was performed on Illumina HiSeq 4000. Demultiplexed reads were mapped using a custom TLA analysis pipeline using the Burrows-Wheeler Aligner mapping software (v 0.7.17) (59). First, reads were mapped to the human genome GRCh38. Then, unaligned sequences were digested in silico with the Nla III restriction site and remapped to the genome. The combined mapping results were used to determine the integration site.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/6/27/eaaz4012/DC1

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. J. A. Beagan, J. E. Phillips-Cremins, On the existence and functionality of topologically associating domains. *Nat. Genet.* **52**, 8–16 (2020).
2. W. A. Bickmore, B. van Steensel, Genome architecture: Domain organization of interphase chromosomes. *Cell* **152**, 1270–1284 (2013).
3. R. Kempfer, A. Pombo, Methods for mapping 3D chromosome architecture. *Nat. Rev. Genet.* **21**, 207–226 (2020).
4. D. M. Ibrahim, S. Mundlos, Three-dimensional chromatin in disease: What holds us together and what drives us apart? *Curr. Opin. Cell Biol.* **64**, 1–9 (2020).
5. V. Dion, J. H. Wilson, Instability and chromatin structure of expanded trinucleotide repeats. *Trends Genet.* **25**, 288–297 (2009).
6. A. L. Castel, J. D. Cleary, C. E. Pearson, Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat. Rev. Mol. Cell Biol.* **11**, 165–170 (2010).
7. D. Kumari, K. Usdin, Chromatin remodeling in the noncoding repeat expansion diseases. *J. Biol. Chem.* **284**, 7413–7417 (2009).
8. Y. H. Fu, D. P. A. Kuhl, A. Pizzuti, M. Pieretti, J. S. Sutcliffe, S. Richards, A. J. M. H. Verkert, J. J. A. Holden, R. G. Fenwick Jr., S. T. Warren, B. A. Oostra, D. L. Nelson, C. T. Caskey, Variation of the CGG repeat at the fragile X site results in genetic instability: Resolution of the Sherman paradox. *Cell* **67**, 1047–1058 (1991).
9. A. J. M. H. Verkerk, M. Pieretti, J. S. Sutcliffe, Y.-H. Fu, D. P. A. Kuhl, A. Pizzuti, O. Reiner, S. Richards, M. F. Victoria, F. Zhang, B. E. Eussen, G.-J. B. van Ommen, L. A. J. Blonden, G. J. Riggins, J. L. Chastain, C. B. Kunst, H. Galjaard, C. Thomas Caskey, D. L. Nelson, B. A. Oostra, S. T. Warran, Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65**, 905–914 (1991).
10. I. Oberlé, F. Rousseau, D. Heitz, C. Kretz, D. Devys, A. Hanauer, J. Boué, M. F. Bertheas, J. L. Mandel, Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science* **252**, 1097–1102 (1991).
11. V. Campuzano, L. Montermini, M. D. Moltò, L. Pianese, M. Cossée, F. Cavalcanti, E. Monros, F. Rodius, F. Duclos, A. Monticelli, F. Zara, J. Cañizares, H. Koutnikova, S. I. Bidichandani, C. Gellera, A. Brice, P. Trouillas, G. De Michele, A. Filla, R. De Frutos, F. Palau, P. I. Patel, S. Di Donato, J.-L. Mandel, S. Cocozza, M. Koenig, M. Pandolfo, Friedreich's ataxia: Autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* **271**, 1423–1427 (1996).
12. D. Kumari, K. Usdin, The distribution of repressive histone modifications on silenced FMR1 alleles provides clues to the mechanism of gene silencing in fragile X syndrome. *Hum. Mol. Genet.* **19**, 4634–4642 (2010).
13. N. Gheldof, T. M. Tabuchi, J. Dekker, The active FMR1 promoter is associated with a large domain of altered chromatin conformation with embedded local histone modifications. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 12463–12468 (2006).

14. R. Pietrobono, E. Tabolacci, F. Zalfa, I. Zito, A. Terracciano, U. Moscato, C. Bagni, B. Oostra, P. Chiurazzi, G. Neri, Molecular dissection of the events leading to inactivation of the FMR1 gene. *Hum. Mol. Genet.* **14**, 267–277 (2005).

15. B. Coffee, F. Zhang, S. Ceman, S. T. Warren, D. Reines, Histone modifications depict an aberrantly heterochromatinized FMR1 gene in fragile X syndrome. *Am. J. Hum. Genet.* **71**, 923–932 (2002).

16. B. Coffee, F. Zhang, S. T. Warren, D. Reines, Acetylated histones are associated with FMR1 in normal but not fragile X–syndrome cells. *Nat. Genet.* **22**, 98–101 (1999).

17. P. K. Chan, R. Torres, C. Yandim, P. P. Law, S. Khadayate, M. Mauri, C. Grosan, N. Chapman-Rothe, P. Giunti, M. Pook, R. Festenstein, Heterochromatinization induced by GAA-repeat hyperexpansion in Friedreich's ataxia can be reduced upon HDAC inhibition by vitamin B3. *Hum. Mol. Genet.* **22**, 2662–2675 (2013).

18. M. V. Evans-Galea, N. Carrodus, S. M. Rowley, L. A. Corben, G. Tai, R. Saffery, J. C. Galati, N. C. Wong, J. M. Craig, D. R. Lynch, S. R. Regner, A. F. D. Brocht, S. L. Perlman, K. O. Bushara, C. M. Gomez, G. R. Wilmot, L. Li, E. Varley, M. B. Delatycki, J. P. Sarsero, FXN methylation predicts expression and clinical outcome in Friedreich ataxia. *Ann. Neurol.* **71**, 487–497 (2012).

19. I. De Biase, Y. K. Chutake, P. M. Rindler, S. I. Bidichandani, Epigenetic silencing in Friedreich ataxia is associated with depletion of CTCF (CCCTC-binding factor) and antisense transcription. *PLOS ONE* **4**, e7914 (2009).

20. S. Al-Mahdawi, R. M. Pinto, O. Ismail, D. Varshney, S. Lymperi, C. Sandi, D. Trabzuni, M. Pook, The Friedreich ataxia GAA repeat expansion mutation induces comparable epigenetic changes in human and transgenic mouse brain and heart tissues. *Hum. Mol. Genet.* **17**, 735–746 (2008).

21. E. Greene, L. Mahishi, A. Entezam, D. Kumari, K. Usdin, Repeat-induced epigenetic changes in intron 1 of the frataxin gene and its consequences in Friedreich ataxia. *Nucleic Acids Res.* **35**, 3383–3390 (2007).

22. J. H. Sun, L. Zhou, D. J. Emerson, S. A. Phyo, K. R. Titus, W. Gong, T. G. Gilgenast, J. A. Beagan, B. L. Davidson, F. Tassone, J. E. Phillips-Cremins, Disease-associated short tandem repeats co-localize with chromatin domain boundaries. *Cell* **175**, 224–238.e15 (2018).

23. A. L. Castel, M. Nakamori, S. Tomé, D. Chitayat, G. Gourdon, C. A. Thornton, C. E. Pearson, Expanded CTG repeat demarcates a boundary for abnormal CpG methylation in myotonic dystrophy patient tissues. *Hum. Mol. Genet.* **20**, 1–15 (2011).

24. D. H. Cho, C. P. Thienes, S. E. Mahoney, E. Analau, G. N. Filippova, S. J. Tapscott, Antisense transcription and heterochromatin at the DM1 CTG repeats are constrained by CTCF. *Mol. Cell* **20**, 483–489 (2005).

25. G. N. Filippova, C. P. Thienes, B. H. Penn, D. H. Cho, Y. J. Hu, J. M. Moore, T. R. Klesert, V. V. Lobanenkov, S. J. Tapscott, CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus. *Nat. Genet.* **28**, 335–343 (2001).

26. J. R. Brouwer, A. Huguet, A. Nicole, A. Munnich, G. Gourdon, Transcriptionally repressive chromatin remodelling and CpG methylation in the presence of expanded CTG-repeats at the DM1 locus. *J. Nucleic Acids* **2013**, 567435 (2013).

27. A. D. Otten, S. J. Tapscott, Triplet repeat expansion in myotonic dystrophy alters the adjacent chromatin structure. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 5465–5469 (1995).

28. X. A. Su, V. Dion, S. M. Gasser, C. H. Freudenreich, Regulation of recombination at yeast nuclear pores controls repair and triplet repeat stability. *Genes Dev.* **29**, 1006–1017 (2015).

29. E. Splinter, E. de Wit, H. J. G. van de Werken, P. Klous, W. de Laat, Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: From fixation to computation. *Methods* **58**, 221–230 (2012).

30. F. A. Klein, T. Pakozdi, S. Anders, Y. Ghavi-Helm, E. E. M. Furlong, W. Huber, FourCSeq: Analysis of 4C sequencing data. *Bioinformatics* **31**, 3085–3091 (2015).

31. R. Raviram, P. P. Rocha, C. L. Müller, E. R. Miraldi, S. Badri, Y. Fu, E. Swanzey, C. Proudhon, V. Snetkova, R. Bonneau, J. A. Skok, 4C-ker: A method to reproducibly identify genome-wide interactions captured by 4C-seq experiments. *PLOS Comput. Biol.* **12**, e1004780 (2016).

32. L. Barbé, S. Lanni, A. López-Castel, S. Franck, C. Spits, K. Keymolen, S. Seneca, S. Tomé, I. Miron, J. Letourneau, M. Liang, S. Choufani, R. Weksberg, M. D. Wilson, Z. Sedlacek, C. Gagnon, Z. Musova, D. Chitayat, P. Shannon, J. Mathieu, K. Sermon, C. E. Pearson, CpG methylation, a parent-of-origin effect for maternal-biased transmission of congenital myotonic dystrophy. *Am. J. Hum. Genet.* **100**, 488–505 (2017).

33. C. Cinesi, L. Aeschbach, B. Yang, V. Dion, Contracting CAG/CTG repeats using the CRISPR-Cas9 nickase. *Nat. Commun.* **7**, 13272 (2016).

34. P. J. P. de Vree, E. de Wit, M. Yilmaz, M. van de Heijning, P. Klous, M. J. A. M. Verstegen, Y. Wan, H. Teunissen, P. H. L. Krijger, G. Geeven, P. P. Eijk, D. Sie, B. Ylstra, L. O. M. Hulsman, M. F. van Dooren, L. J. C. M. van Zutven, A. van den Ouweland, S. Verbeek, K. W. Van Dijk, M. Cornelissen, A. T. Das, B. Berkhout, B. Sikkema-Raddatz, E. van den Berg, P. van der Vlies, D. Weening, J. T. den Dunnen, M. Matusiak, M. Lamkanfi, M. J. L. Ligtenberg, P. ter Brugge, J. Jonkers, J. A. Foekens, J. W. Martens, R. van der Luijt, H. K. P. van Amstel, M. van Min, E. Splinter, W. de Laat, Targeted sequencing by proximity ligation for comprehensive variant detection and local haplotyping. *Nat. Biotechnol.* **32**, 1019–1025 (2014).

35. A. Rada-Iglesias, F. G. Grosveld, A. Papantonis, Forces driving the three-dimensional folding of eukaryotic genomes. *Mol. Syst. Biol.* **14**, e8214 (2018).

36. M. Merkenschlager, E. P. Nora, CTCF and cohesin in genome folding and transcriptional gene regulation. *Annu. Rev. Genomics Hum. Genet.* **17**, 17–43 (2016).

37. R. Adihe Lokanga, X.-N. Zhao, A. Entezam, K. Usdin, X inactivation plays a major role in the gender bias in somatic expansion in a mouse model of the fragile X–related disorders: Implications for the mechanism of repeat expansion. *Hum. Mol. Genet.* **23**, 4985–4994 (2014).

38. D. Wöhrle, U. Salat, H. Hameister, W. Vogel, P. Steinbach, Demethylation, reactivation, and destabilization of human fragile X full-mutation alleles in mouse embryocarcinoma cells. *Am. J. Hum. Genet.* **69**, 504–515 (2001).

39. G. De Michele, F. Cavalcanti, C. Criscuolo, L. Pianese, A. Monticelli, A. Filla, S. Cocozza, Parental gender, age at birth and expansion length influence GAA repeat intergenerational instability in the X25 gene: Pedigree studies and analysis of sperm from patients with Friedreich's ataxia. *Hum. Mol. Genet.* **7**, 1901–1906 (1998).

40. Y. Ghavi-Helm, A. Jankowski, S. Meiers, R. R. Viales, J. O. Korbel, E. E. M. Furlong, Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat. Genet.* **51**, 1272–1282 (2019).

41. A. R. Barutcu, P. G. Maass, J. P. Lewandowski, C. L. Weiner, J. L. Rinn, A TAD boundary is preserved upon deletion of the CTCF-rich Firre locus. *Nat. Commun.* **9**, 1444 (2018).

42. E. Rodríguez-Carballo, L. Lopez-Delisle, Y. Zhan, P. J. Fabre, L. Beccari, I. El-Idrissi, T. H. Nguyen Huynh, H. Ozadam, J. Dekker, D. Duboule, The *HoxD* cluster is a dynamic and resilient TAD boundary controlling the segregation of antagonistic regulatory landscapes. *Genes Dev.* **31**, 2264–2281 (2017).

43. V. Dion, Y. Lin, L. Hubert Jr., R. A. Waterland, J. H. Wilson, Dnmt1 deficiency promotes CAG repeat expansion in the mouse germline. *Hum. Mol. Genet.* **17**, 1306–1317 (2008).

44. D. Gaidatzis, A. Lerch, F. Hahne, M. B. Stadler, QuasR: Quantification and annotation of short reads in R. *Bioinformatics* **31**, 1130–1132 (2015).

45. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

46. P. V. Kharchenko, M. Y. Tolstorukov, P. J. Park, Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat. Biotechnol.* **26**, 1351–1359 (2008).

47. Q. Li, J. B. Brown, H. Huang, P. J. Bickel, Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* **5**, 1752–1779 (2011).

48. C. S. Ross-Innes, R. Stark, A. E. Teschendorff, K. A. Holmes, H. R. Ali, M. J. Dunning, G. D. Brown, O. Gojis, I. O. Ellis, A. R. Green, S. Ali, S.-F. Chin, C. Palmieri, C. Caldas, J. S. Carroll, Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389–393 (2012).

49. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

50. F. P. A. David, J. Delafontaine, S. Carat, F. J. Ross, G. Lefebvre, Y. Jarosz, L. Sinclair, D. Noordermeer, J. Rougemont, M. Leleu, HTSstation: A web application and open-access libraries for high-throughput sequencing data analysis. *PLOS ONE* **9**, e85879 (2014).

51. D. Noordermeer, M. Leleu, E. Splinter, J. Rougemont, W. De Laat, D. Duboule, The dynamic architecture of Hox gene clusters. *Science* **334**, 222–225 (2011).

52. C. Walter, D. Schuetzmann, F. Rosenbauer, M. Dugas, Benchmarking of 4C-seq pipelines based on real and simulated data. *Bioinformatics* **35**, 4938–4945 (2019).

53. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

54. H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).

55. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.J.* **17**, 10–12 (2011).

56. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

57. D. W. Huang, B. T. Sherman, R. A. Lempicki, Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).

58. D. W. Huang, B. T. Sherman, R. A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).

59. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).

60. H. Yao, K. Brick, Y. Evrard, T. Xiao, R. D. Camerini-Otero, G. Felsenfeld, Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. *Genes Dev.* **24**, 2543–2555 (2010).

61. H. J. G. van de Werken, G. Landan, S. J. B. Holwerda, M. Hoichman, P. Klous, R. Chachik, E. Splinter, C. Valdes-Quezada, Y. Oz, B. A. M. Bouwman, M. J. A. M. Verstegen, E. de Wit, A. Tanay, W. de Laat, Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat. Methods* **9**, 969–972 (2012).

Citation: G. A. Ruiz Buendía, M. Leleu, F. Marzetta, L. Vanzan, J. Y. Tan, V. Ythier, E. L. Randall, A. C. Marques, T. Baubec, R. Murr, I. Xenarios, V. Dion, Three-dimensional chromatin interactions remain stable upon CAG/CTG repeat expansion. *Sci. Adv.* **6**, eaaz4012 (2020).

Science Advances

Three-dimensional chromatin interactions remain stable upon CAG/CTG repeat expansion

Gustavo A. Ruiz Buendía, Marion Leleu, Flavia Marzetta, Ludovica Vanzan, Jennifer Y. Tan, Victor Ythier, Emma L. Randall, Ana C. Marques, Tuncay Baubec, Rabih Murr, Ioannis Xenarios and Vincent Dion

ARTICLE TOOLS	http://advances.sciencemag.org/content/6/27/eaaz4012
SUPPLEMENTARY MATERIALS	http://advances.sciencemag.org/content/suppl/2020/06/29/6.27.eaaz4012.DC1
REFERENCES	This article cites 61 articles, 11 of which you can access for free http://advances.sciencemag.org/content/6/27/eaaz4012#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service