Stem Cell Reports
Article
OPEN ACCESS

# Uncovering the Role of Hypermethylation by CTG Expansion in Myotonic Dystrophy Type 1 Using Mutant Human Embryonic Stem Cells

Shira Yanovsky-Dagan,[1] Michal Avitzour,[1] Gheona Altarescu,[2] Paul Renbaum,[2] Talia Eldar-Geva,[3] Oshrat Schonberger,[3] Stella Mitrani-Rosenbaum,[4] Ephrat Levy-Lahad,[2] Ramon Y. Birnbaum,[5] Lior Gepstein,[6] Silvina Epsztejn-Litman,[1] and Rachel Eiges[1,*]
[1]Stem Cell Research Laboratory, Medical Genetics Institute
[2]Zohar PGD Lab, Medical Genetics Institute
[3]IVF Unit
Shaare Zedek Medical Center affiliated with the Hebrew University School of Medicine, Jerusalem 91031, Israel
[4]Goldyne Savad Institute for Gene Therapy, Hadassah Hebrew University Medical Center, Jerusalem 91240, Israel
[5]Department of Life Sciences, Ben-Gurion University of the Negev, Beer Sheva 84105, Israel
[6]Sohnis Family Research Laboratory for Cardiac Electrophysiology and Regenerative Medicine, Bruce Rappaport Faculty of Medicine, Technion-Israel Institute of Technology, Haifa 31096, Israel
*Correspondence: rachela@szmc.org.il
http://dx.doi.org/10.1016/j.stemcr.2015.06.003


## SUMMARY

CTG repeat expansion in *DMPK*, the cause of myotonic dystrophy type 1 (DM1), frequently results in hypermethylation and reduced *SIX5* expression. The contribution of hypermethylation to disease pathogenesis and the precise mechanism by which *SIX5* expression is reduced are unknown. Using 14 different DM1-affected human embryonic stem cell (hESC) lines, we characterized a differentially methylated region (DMR) near the CTGs. This DMR undergoes hypermethylation as a function of expansion size in a way that is specific to undifferentiated cells and is associated with reduced *SIX5* expression. Using functional assays, we provide evidence for regulatory activity of the DMR, which is lost by hypermethylation and may contribute to DM1 pathogenesis by causing *SIX5* haplo-insufficiency. This study highlights the power of hESCs in disease modeling and describes a DMR that functions both as an exon coding sequence and as a regulatory element whose activity is epigenetically hampered by a heritable mutation.

## INTRODUCTION

Myotonic dystrophy type 1 (DM1) is an autosomal dominant muscular dystrophy that affects a wide range of body systems (DM1 [OMIM: 160900]). It results from a trinucleotide CTG repeat expansion (50–4,000 copies) in the 3′ UTR of the dystrophia myotonica protein kinase gene (*DMPK*) (Aslanidis et al., 1992; Brook et al., 1992). The CTG repeat region, which resides within a CpG island (CGI), commonly results in hypermethylation and the spread of heterochromatin when expanded (Cho et al., 2005; Filippova et al., 2001; Steinbach et al., 1998). Hypermethylation is largely age- and tissue-specific and does not necessarily correlate with expansion size in somatic cells of patients (López Castel et al., 2011; Spits et al., 2010). In addition, when the CTG repeats expand, they commonly result in a reduction in the expression of a neighboring gene, *SIX5* (Klesert et al., 1997, 2000; Korade-Mirnics et al., 1999; Sarkar et al., 2000, 2004; Thornton et al., 1997). The contribution of hypermethylation to disease pathogenesis is still not fully understood, nor is the precise mechanism by which CTG expansion leads to *SIX5* reduction in *cis*. Using a wide range of DM1-affected human embryonic stem cell (hESC) lines, we aimed to uncover the mechanistic relationship between CTG expansion, aberrant methylation, and reduced expression of *SIX5* in DM1.

## RESULTS

### Derivation and Characterization of DM1 hESCs

Fourteen different mutant hESC lines were established from DM1-affected preimplantation embryos. This exclusive set of DM1-affected cell lines, which displays the typical characteristics of hESCs (Figure S1), represents a wide range of maternally and paternally inherited expansions, bearing from 180 to more than 2,000 CTG repeats (Table 1).

### Characterization of a Disease-Associated, Differentially Methylated Region Upstream of the CTG Repeats in DM1 hESCs

To assess whether normal and expanded alleles differ in their DNA methylation patterns in undifferentiated cells, we employed a methylation-sensitive Southern blot assay that relies on the digestion of a SacI-HindIII fragment with either MspI or its methylation-sensitive isoschizomer, HpaII. Because the SacI-HindIII fragment contains 26 MspI/HpaII recognition sites, of which only one is located downstream of the repeats (Figures 1A and 1B), the digestion of this segment with either HpaII or MspI facilitates the identification of methylation upstream of the CTGs. Using this test on wild-type (WT) and affected hESCs, we show that abnormal methylation is already established in

**Table 1. DM1-Affected hESC Line Collection**

| HESC Line | Family | Parental Origin of Expansion | CTG Copy Number | SNP1 | SNP2 | SNP3 | SNP4 |
|---|---|---|---|---|---|---|---|
| SZ-DM1 | A | paternal | 1000 | G/**T** | | C/**G** | C/**G** |
| SZ-DM2 | A | paternal | 430,500,1000 | G/**T** | G/**T** | | C/**G** |
| SZ-DM3 | B | maternal | 2000 | | | | C/**G** |
| SZ-DM4 | C | paternal | 180 | | G/**T** | | C/**G** |
| SZ-DM5 | D | paternal | 300 | G/**T** | G/**T** | | C/**G** |
| SZ-DM6 | A | paternal | 1060 | G/**T** | G/**T** | C/**G** | C/**G** |
| SZ-DM7 | E | paternal | 300 | G/**T** | G/**T** | | C/**G** |
| SZ-DM8 | B | maternal | 400,1900,2000 | | | | C/**G** |
| SZ-DM9 | F | paternal | 180 | | G/**T** | C/**G** | C/**G** |
| SZ-DM11 | A | paternal | 500 | | | | C/**G** |
| SZ-DM12 | A | paternal | 630 | | | | C/**G** |
| SZ-DM14 | H | maternal | 2000 | | | | C/**G** |
| SZ-DM20 | A | paternal | 300 | G/**T** | G/**T** | C/**G** | C/**G** |
| LIS-DM[a] | G | maternal | 600–1600 | | | | C/**G** |

Information related to the newly established DM1 hESC lines, including family code, parental origin of expansion, CTG repeat copy number, and informative SNPs. SNP1 (rs915915) and SNP2 (rs635299) are intronic polymorphisms located 1.5 kb and 650 bp upstream of the repeats, respectively. SNP3 (rs527221) and SNP4 (rs2014377) are exonic polymorphisms located in exon 10 and exon 3 of the *DMPK* and *SIX5* genes, respectively. Alleles that were confirmed to be in linkage disequilibrium with the CTG expansion by haplotype analysis are marked in bold. See also Figure S1 for a full characterization of DM1 hESC lines.
[a]Provided by the Racine IVF Unit, Tel-Aviv Sourasky Medical Center.

the undifferentiated state and that it is exclusively acquired by expansions greater than 300 CTG repeat copies (Figure 1C; Figure S2A). In addition, we find that, in hESCs, a clear association exists between expansion size and extent of methylation. That is, the larger the expansion, the larger the region of methylated DNA. Interestingly, the 1.3- and 1.6-kb bands (arrows in Figures 1C and S2A) indicate that the sites adjacent to the CTG repeats only become methylated in the larger expansions, attesting to a distinct pattern of acquisition of methylation in expanded alleles.

To understand the methylation events in expanded alleles at a higher resolution, we first defined the 5′ border of the differentially methylated region (DMR) by bisulfite colony sequencing in WT hESCs and found that the DMR only begins 700 base pairs (bp) into the CGI, 900 bp upstream of the CTGs (intron 13 of *DMPK*) (Figure S3). We then determined methylation levels in region E of the DMR in DM1 hESC lines by allele-specific bisulfite sequencing, utilizing an informative SNP that resides within this region. Allelic association between the informative SNP and the expansion was carried out by haplotype analysis using parental DNAs in each cell line. Using this approach, we found that repeat expansion size strongly correlates with hypermethylation of this region (650 bp upstream from the repeats), ranging from 0% (<300 CTGs) to over 88% (1,000 CTGs) ($\rho = 0.94114$; Spearman's rank correlation, $p < 0.01$) (Figures 2A and 2B; Figure S4A). Absolute methylation levels in cell lines with no informative SNPs in this region further corroborated these results (Figure S4B). Furthermore, statistical validation showed that low levels of abnormal methylation can already be observed in expansions of 300 CTG copies as opposed to normal alleles in informative cell lines, confirming the Southern blot results ($p < 0.01$ by paired t test; Figure 2C). Taken together with the Southern blot analysis (which indicated that the last region to undergo methylation was the region adjacent to the CTG repeats), these results indicate that abnormal methylation of the DMR begins from the 5′ end of the CGI (already methylated) and spreads toward the CTG repeats.

To further substantiate our findings, we explored whether reprogramming somatic cells to induced pluripotent stem cells (iPSCs) would recapitulate aberrant methylation patterns specific to undifferentiated cells. Primary fibroblasts with a CTG expansion of approximately 2,000 repeats, which were only partly methylated upstream of the repeats, were reprogrammed according to Takahashi et al. (2007). This resulted in the establishment of three

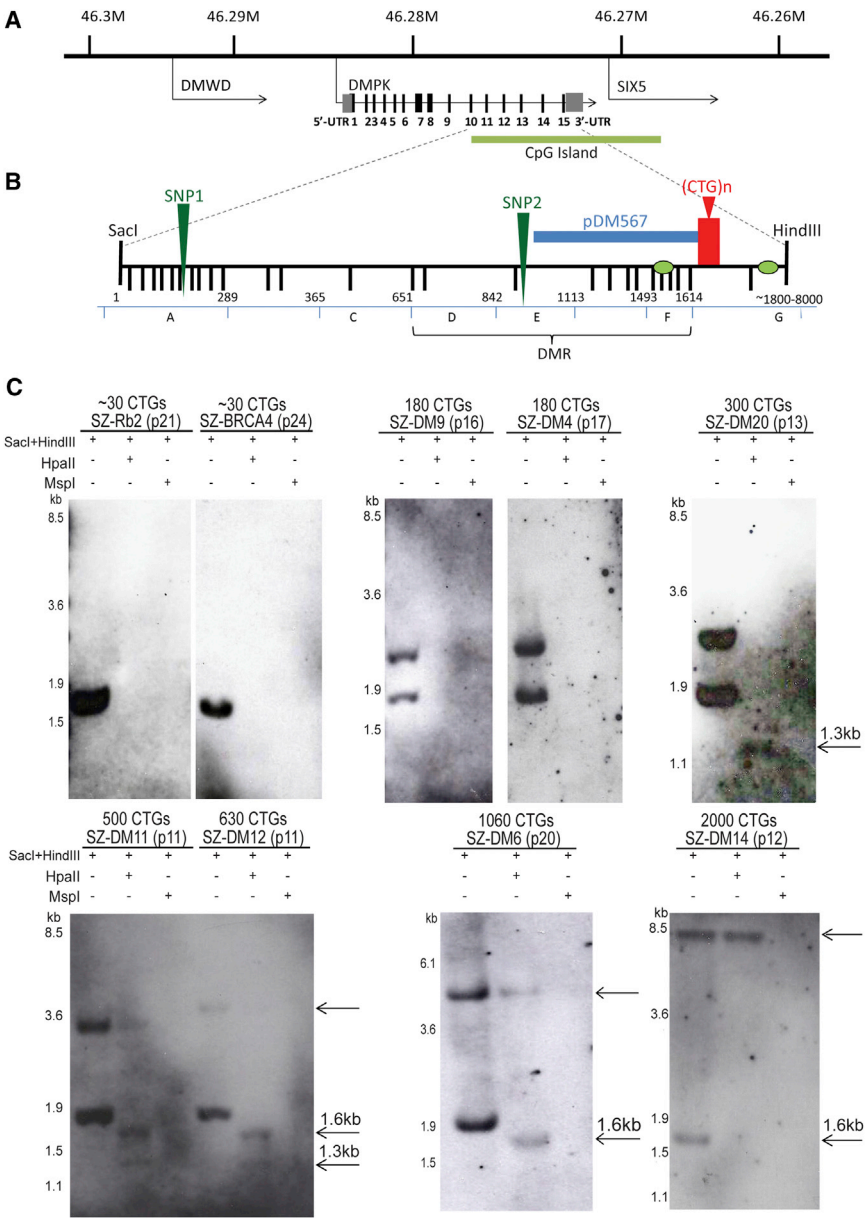



**Figure 1. DNA Methylation Analysis Upstream of the CTGs by Southern Blot Analysis**

(A) Schematic illustration of the human DM1 locus.

(B) Detailed description of the 3′ portion of the DM1 locus, spanning from exon 11 to the CTG repeats (~1.8 kb long in WT alleles [~20 CTGs] and up to 8 kb long in expanded alleles). Twenty-six HpaII/MspI recognition sites, embedded within the SacI-HindIII restriction fragment, are indicated by black vertical bars. Also indicated are the probe used for the Southern blot analysis (pDM567), two polymorphic sites (SNP1 and SNP2), six proximal genomic regions analyzed in this study by bisulfite sequencing/pyrosequencing (A, C, D, E, F, and G), region D–F representing the DMR, and two putative CTCF binding sites upstream and downstream of the repeats (green ellipses). The numbers under genomic regions A, C, D, E, F, and G indicate the relative distance from the SacI restriction site (position 1 corresponds with chr19:46,275,123 on the hg19 human genome assembly).

(C) Representative Southern blots from WT (SZ-RB2 and SZ-BRCA4) of, altogether, seven WT hESC lines examined and DM1-affected hESC lines with small (<300 CTGs, SZ-DM9 and SZ-DM4), intermediate (~300 CTGs, SZ-DM20), moderate to large (450–650 CTGs, SZ-DM11 and SZ-DM12), and particularly large expansions (>650 CTGs, SZ-DM6 and SZ-DM14). To determine CTG repeat number by Southern blot analysis, genomic DNA was double-digested with SacI and HindIII and then hybridized with the pDM567 probe, resulting in an ~1.8-kb (normal allele with 20 CTGs) and a larger fragment (expanded allele) (left lane in each panel). To investigate aberrant methylation upstream of the repeats, the SacI-HindIII fragments were further processed by parallel restriction with MspI (+MspI) and its methylation-sensitive isoschizomer, HpaII (+HpaII). Marked by arrows are methylated 1.3-kb, 1.6-kb, and larger fragments. The CTG repeat number and passage number (in parentheses) are indicated for each cell line at the top of each blot.

See also Figure S2 for additional Southern blots of DM1 hESC lines and Figure S3 for identification of the CpG Island 5′ methylation border in WT hESCs.

iPSC clones (iPS-4, iPS-19, and iPS-24), each featuring typical characteristics of pluripotent stem cells and a normal karyotype (Figure S1). Using methylation-sensitive Southern blot analysis, we found that reprogramming led to an appreciable gain of abnormal methylation in all iPS cell clones but no change in expansion size (Figure S2B). In addition, the bisulfite pyro-sequencing results indicated a marked increase in methylation levels of the mutant allele immediately upstream of the CTG repeats in DM1-reprogrammed iPS cells relative to parental fibroblasts (region F; Figure S5E, bottom). Therefore, our results suggest that aberrant methylation patterns are unique to undifferentiated DM1-affected cells and can be re-established following cell reprogramming.

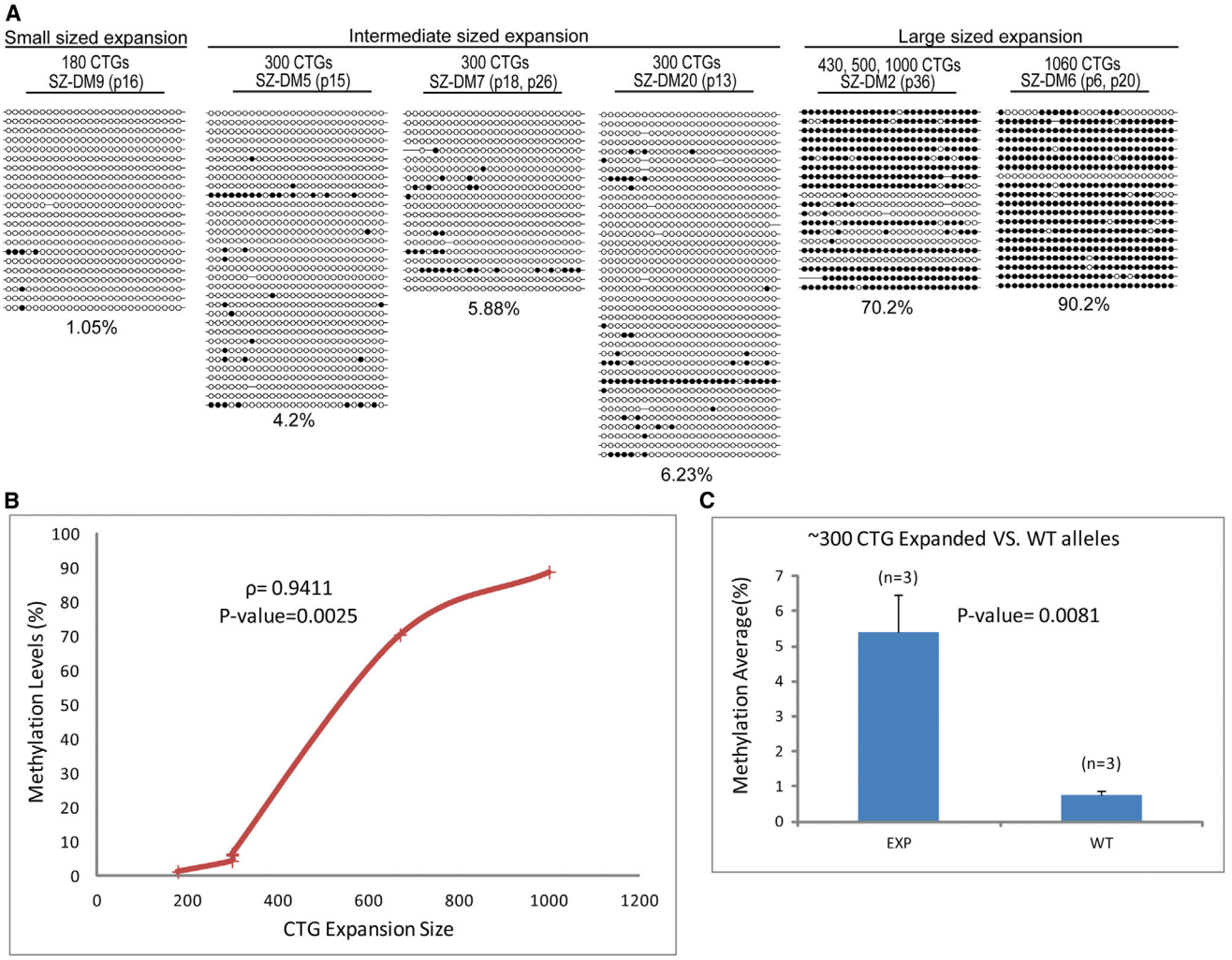



**Figure 2. Allele-Specific Methylation Levels in the DMR Correlates with Expansion Size in DM1 hESCs**

(A) Allele-specific bisulfite sequencing of the expanded allele within the DMR (region E, 488–777 bp upstream of the repeats, overlapping the SNP2 site). Full circles correspond to methylated CpGs, whereas empty circles represent unmethylated CpGs. For each cell line, methylation levels (%) were determined by the analysis of at least 20 molecules by bisulfite sequence. Methylation levels (%), passage number, and CTG expansion size are indicated for each cell line.

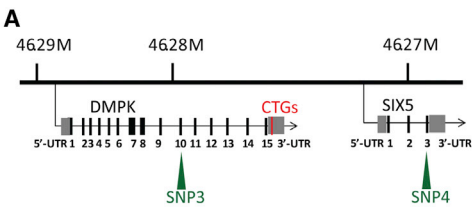
(B) Graphical summary of the results in (A), showing the significant correlation that exists between expansion size and methylation levels ($\rho = 0.94114$; Spearman's rank correlation test, $p < 0.01$).

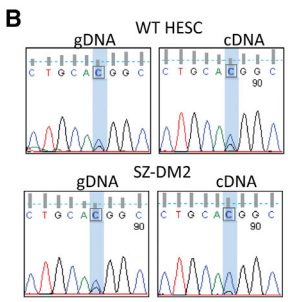
(C) Intermediate-sized expansions (~300 CTG repeats) of SZ-DM5, SZ-DM7, and SZ-DM20 samples exhibit significantly higher methylation levels than their WT counterpart alleles ($p < 0.01$ by paired t test). n, number of cell lines examined. Error bars represent SD.

See also Figure S3 for identification of the CpG Island 5′ methylation border in WT hESCs and Figure S4 for allele-specific bisulfite sequencing of the WT allele in informative DM1 hESC lines, and for absolute bisulfite sequencing within non-informative DM1 hESC lines and iPSC clones.

## Association between Aberrant Methylation and Local Gene Transcription

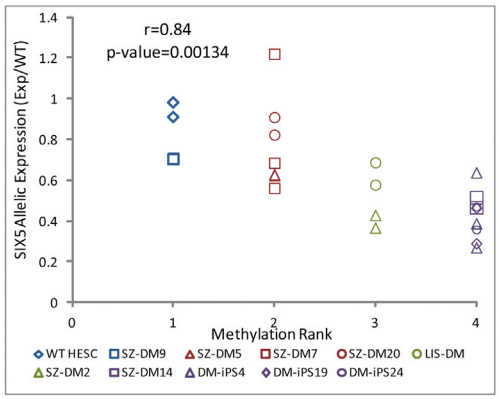
To link abnormal methylation with disease pathogenesis, we identified informative SNPs within exon 10 of *DMPK* (SNP3) and exon 3 of *SIX5* (SNP4) and associated gene expression with mutated alleles in DM1 hESCs (Figure 3A; Table 1). In parallel, we monitored allele-specific alterations in *SIX5* expression in DM1 iPSCs that carried a polymorphism in *SIX5* but not in *DMPK*. We utilized these polymorphisms to employ a commonly used method that relies on pair-wise comparisons between genomic DNA and cDNA sequences to identify allelic expression imbalances (AEIs) (Ben-David et al., 2011; Figure 3B). Our analysis identified a striking inverse correlation ($\rho = 0.84$; Spearman's rank correlation coefficient, $p = 0.00134$) between the methylation level in the E region of the DMR and *SIX5* expression




C

| Methylation (rank) | Cell Line | SIX5 Expression (Exp/WT) |
|---|---|---|
| 0% (1) | WT HESC | 0.94± 0.05 |
| | SZ-DM9 | 0.71± 0.005 |
| ~5% (2) | SZ-DM5 | 0.63± 0.002 |
| | SZ-DM7 | 0.74± 0.35 |
| | SZ-DM20 | 0.87± 0.061 |
| 70-80% (3) | LIS-DM | 0.63± 0.077 |
| | SZ-DM2 | 0.4± 0.044 |
| 100% (4) | SZ-DM14 | 0.49± 0.035 |
| | DM-iPS-4 | 0.38± 0.188 |
| | DM-iPS-19 | 0.36± 0.122 |
| | DM-iPS-24 | 0.41± 0.08 |




E

| Methylation in E Region (%) | | | | | | | | | | | | | AVE (%) | *SIX5* expression (Exp/WT) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| WT HESC-Cardio | | | | | | | | | | | | | 1.4 | 1.01± 0.05 |
| SZ-DM14-Cardio | | | | | | | | | | | | | 37.6 | 0.27± 0.02 |
| CpG position | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | | |


**Figure 3. Association between Aberrant Methylation and *SIX5* Gene Transcription**

(A) Schematic illustration of the human DM1 locus, including the *DMPK* gene and its adjacent neighbor gene, *SIX5*. SNP3 and SNP4 are located in the coding regions of *DMPK* and *SIX5*, respectively.

(B) Identification of *SIX5* AEIs. SNP4 sequencing results (shaded) in representative genomic DNA (gDNA) and cDNA samples from WT and DM1-affected hESCs (SZ-RB26 and SZ-DM2, respectively). The expression ratio between the alleles in each cell line was estimated using PeakPicker2 software.

(C) Cell line-specific comparison of expanded allele methylation levels in region E (as determined by allele-specific and absolute bisulfite sequencing) with the mean *SIX5* expression ratio of expanded allele/WT allele. The assayed cell lines, WT (SZ-RB26) and DM1-affected hESCs, were arbitrarily ranked into groups according to the levels of aberrant methylation in region E. *SIX5* expression by the expanded allele correlates well with methylation levels (region E) in 11 pluripotent cell lines ($\rho = 0.84$; Spearman's rank correlation, $p < 0.01$). The correlation was found to be statistically significant ($p = 0.001$) according to a permutation test (50,000 permutations). Values of allele-specific transcription of *SIX5* were determined based on the analysis of two to three biological replicas for each cell line. SD was calculated for each cell line.

(D) Graphical summary of the tabulated results in (C).

(E) Absolute methylation levels in region E, as determined by bisulfite pyrosequencing, and the *SIX5* allelic expression ratio in in vitro-differentiated cardiomyocytes from WT (SZ-RB26, n = 2) and SZ-DM14 (n = 4) hESCs ($p < 0.01$ by paired t test). The 13 CpG sites tested are indicated by shaded squares according to methylation levels (see colored legend). Mean absolute methylation levels across all tested CpG positions are indicated to the right in each line.

See also Figure S5 for *DMPK* expression, methylation downstream to the CTGs, expression of cardiospecific markers, and CTCF binding analysis.

(Figures 3C and 3D). That is, the higher the levels of methylation upstream from the repeats, the lower the *SIX5* expression levels that were assayed. Importantly, no correlation was found between hypermethylation and allele-specific alterations in *DMPK* expression (Figure S5A). Furthermore, we could not find an association between the reduction in *SIX5* expression and methylation downstream of the CTGs (region G) or at the promoter region of *SIX5* (Figure S5B), regardless of expansion size. Therefore, we carried on with our study, focusing on the DMR region.

To characterize the epigenetic status of the DMR in the context of myotonic dystrophy disease symptoms, we generated functional cardiomyocytes by in vitro differentiation, taking into account the frequent involvement of cardiac complications in DM1 patients (Lund et al., 2014; Martorell et al., 1997; Sovari et al., 2007). Using an optimized protocol for cardiac differentiation (Burridge et al., 2014), we established a large number of contracting cardiomyocytes from WT and DM1 hESCs that express cardiac-specific markers, as confirmed by RT-PCR (Figure S5C). We examined DMR hypermethylation and allele-specific reduction in *SIX5* expression in the WT and DM1 in vitro-differentiated cardiomyocytes. Notably, DMR methylation levels (E region) were much higher in DM1-affected cardiomyocytes compared with the WT control cells, and this was accompanied by reduced expression of *SIX5* from the expanded allele (30% of the WT allele) (Figure 3E). Therefore, it appears that increased

methylation upstream of the CTG repeats, in the coding region of *DMPK*, hampers the expression of the downstream proximal gene, *SIX5*, in DM1 hESCs and in in vitro-differentiated cardiomyocytes.

### Functional Association between DMR Hypermethylation and the Reduction in *SIX5* Expression

To address the mechanistic association between aberrant methylation levels and the decrease in *SIX5* expression, we first ruled out the previously suggested hypothesis that *SIX5* expression is regulated by the insulator binding protein CTCF (Filippova, 2001, 2008). CTCF is displaced when its recognition site is hypermethylated (Filippova, 2008). Using chromatin immunoprecipitation (ChIP) analysis, we first confirmed the occupancy of CTCF to its binding site upstream of the CTG repeats in both WT and affected hESCs (Figure S5D). No CTCF enrichments were found downstream of the CTGs (Figure S5D). Next, by bisulfite pyro-sequencing of the bound and input fractions in the CTCF binding site (region F; Figure 1B), we validated CTCF differential binding affinity to its unmethylated recognition site upstream of the CTG repeats (Figure S5E, top). Then we searched for a correlation between hypermethylation of the CTCF binding site, as determined by bisulfite pyro-sequencing, and allele-specific reduction of *SIX5* mRNA in the affected cell lines (Figure 3D). However, no correlation was found (Figure S5E, bottom). Therefore, we ruled out the possibility that loss of CTCF binding contributed to the reduction in *SIX5* expression by hypermethylation, as proposed previously (Filippova, 2001, 2008).

Next we asked whether hypermethylation by CTG expansion hampers the activity of a putative *SIX5* regulatory element within the *DMPK* gene. Using both in vivo and in vitro functional assays, we explored the regulatory role of the DMR sequence.

Using a zebrafish functional assay, we explored whether the DMR sequence could function as an active regulatory element in vivo. Despite the lack of a *SIX5* homolog in zebrafish, we show that the DMR sequence alone (Figure 4A, fragment I) drives GFP expression. This is in contrast to the other fragments (*SIX5* promoter and the entire region; Figure 4A, fragments II and III), which failed to drive GFP expression. GFP expression under the regulation of the DMR, which, although weak, was restricted to the heads of transgenic animals and was predominantly expressed in the epidermis (beneath the eye), hindbrain, and branchial arches at 48 and 72 hr post-fertilization (hpf) (Figure 4B).

We then examined whether the DMR could drive luciferase reporter gene (luc+) expression by transient transfection in three different cell lines (293T, COS-7, and HepG2). We show that, although the empty vector completely lacks reporter gene activity, insertion of the DMR alone (Figure 4A, fragment I) results in a 10- to 15-fold increase over basal activity obtained either by the insertion of the *SIX5* promoter (Figure 4A, fragment II), the entire region spanning from the DMR to the *SIX5* promoter (Figure 4A, fragment III), or the DMR in an opposite orientation. The basal activity obtained by the entire region (extending from the DMR to the *SIX5* promoter) suggests an inhibitory effect of a downstream sequence that is not apparent in vivo. Together, these transient expression experiments suggest that the DMR alone is capable of strongly activating reporter gene expression in vitro when in a proper orientation and independent of the *SIX5* promoter (Figure 4C).

To examine the effects of methylation on the regulatory activity of this region, we in vitro-methylated the DMR vector using the CpG methyltransferase M.SssI and used it in the transfection assay (validated by digestion with the restriction endonuclease HpaII, which is sensitive to methylation; data not shown). Strikingly, methylation of the construct led to complete silencing of reporter gene expression in all cell lines (Figure 4C), indicating that DNA methylation of the DMR results in the loss of its ability to stimulate transcription.

To demonstrate a possible regulatory role for the DMR in a disease-relevant tissue, we examined whether the DMR alone is sufficient to drive the expression of an eGFP reporter in in vitro-differentiated cardiomyocytes. Undifferentiated DMR-eGFP-transfected hESC clones were generated and propagated, resulting in neomycin-resistant colonies with a variegated expression of eGFP. High eGFP-expressing hESCs were isolated by fluorescence-activated cell sorting (FACS) and induced to differentiate into contracting cardiomyocytes. This approach allowed us to successfully generate DMR-driven fluorescent green-emitting, functioning cardiomyocytes. Taken together with the observation that the regulatory activity of the DMR is methylation dependent (Figure 4C) and that an inverse correlation between DMR hypermethylation and SIX5 expression exists in differentiated cardiomyocytes (Figure 3E), we confirmed the potential role of the DMR as a distal regulatory element in a disease-relevant cell type (Figure 4D; Movies S1, S2, S3, and S4).

Together, we provide compelling evidence for the discovery of an epigenetically controlled regulatory element within the coding sequence of *DMPK* that is capable of driving gene expression in transgenic animals as well as in various cell lines/types, including cardiac muscle cells. Its regulatory activity most likely affects *SIX5* expression and is hampered by the gain of aberrant methylation.

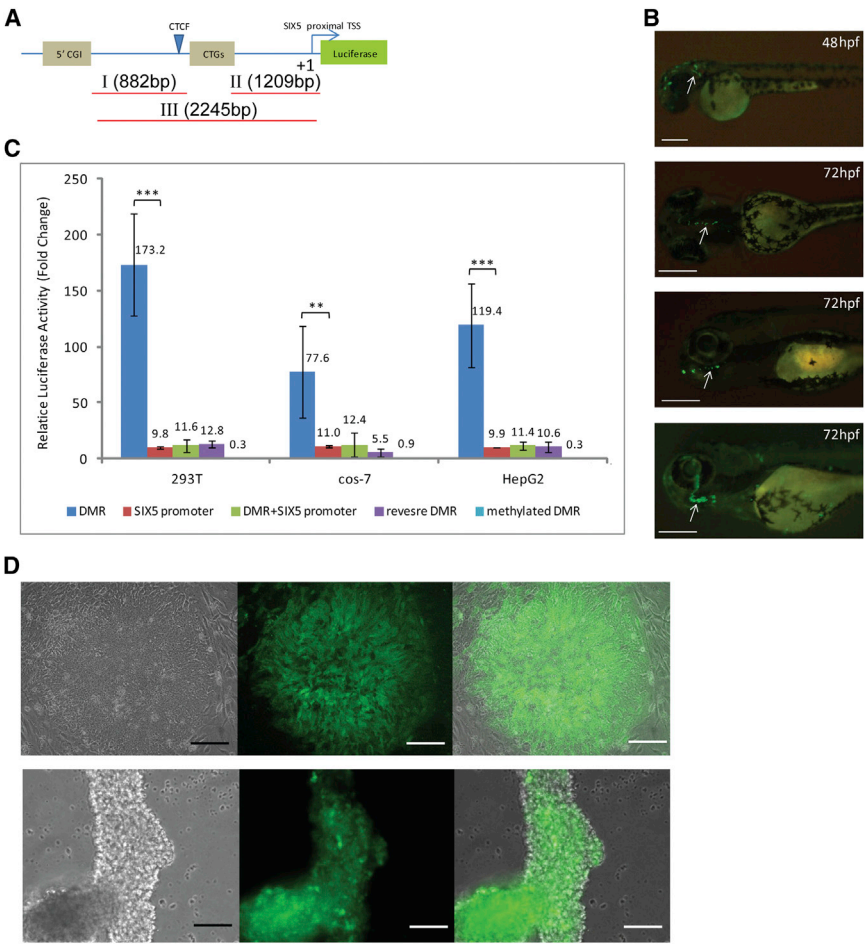



**Figure 4. Functional Assay for the Regulatory Role of the DMR**

(A) Schematic illustration describing the different inserts tested in this study. Fragments I, II, and III designate the DMR sequence in both orientations, the *SIX5* promoter, and the entire region that spans from the DMR to the *SIX5* promoter (includes 12 repeats), respectively. The size of each fragment is indicated in parentheses.

(B) Zebrafish enhancer assay for the DMR region. GFP expression was observed in the epidermis and branchial arches of zebrafish embryos (marked by arrows) at 48 and 72 hr post-fertilization (hpf). Scale bars, 100 μm.

(C) Relative luciferase activity from transiently transfected 293T, COS-7, and HepG2 cells. Luciferase activity was measured 48 hr after transfection. *Renilla* expression levels were used to normalize cell number, transfection efficiency, and general effects on gene transcription. Data were normalized to luciferase levels of the empty vector pGL3 Basic. Bar graphs stand for mean values (indicated on top of each bar) and SD from three independent experiments with three technical replicates each (paired t test, $^{**}p < 0.01$, $^{***}p < 0.001$).

(D) EGFP expression driven by the DMR in stably transfected undifferentiated hESCs (top) and in in vitro-differentiated contracting cardiomyocytes derived from them (bottom) (magnification, 200×; scale bars, 100 μm).

See also Movies S1, S2, S3, and S4 presenting fluorescent green-emitting functioning cardiomyocytes driven by the DMR.

## DISCUSSION

Here we report on 14 different mutant hESC lines, termed DM1 hESCs, that were derived from DM1-affected preimplantation embryos. Previous studies employed DM1 hESCs to explore the causal role of CTG expansion on neurite formation and neuromuscular connections and for characterizing the somatic instability of the CTG repeats in undifferentiated cells (De Temmerman et al., 2008; Marteyn et al., 2011; Seriola et al., 2011). However, neither of these studies addressed the consequences of CTG expansion on DNA methylation and how this change may contribute to disease pathogenesis. In addition, a comprehensive study characterizing aberrant methylation in a wide range of tissues from DM1 aborted fetuses and postmortem samples demonstrates that hypermethylation is largely tissue- and age-specific but not necessarily correlated with expansion size, making it difficult to assign a function to abnormal methylation in disease pathogenesis (López Castel et al., 2011).

Taking advantage of a large set of DM1 hESC lines comprised of maternally and paternally inherited expansions bearing from 180 to more than 2,000 CTG repeats, we finely characterized a disease-associated DMR, 650 bp upstream to the CTG repeats, that abnormally gains methylation in a way that strongly correlates with expansion size ($\rho = 0.94114$; Spearman's rank correlation, $p < 0.01$). We show that this association, which is restricted to undifferentiated cells, is triggered by expansions of more than 300 CTG repeats and is tightly linked with allele-specific reduction in *SIX5* (but not *DMPK*) transcription ($\rho = 0.84$; Spearman's rank correlation coefficient, $p = 0.00134$), as determined by allele expression imbalances. Furthermore, we show that the relationship between expansion size, methylation extent, and allele reduction in *SIX5* is reproduced in patient-derived iPSCs following cell

reprogramming and is maintained in in vitro-differentiated cardiomyocytes, a disease-relevant cell type. Together, our findings in DM1 hESCs and patient-derived iPSCs correspond with the decline in *SIX5* expression in DM1 patients and with the expression of various aspects of DM1 pathology in *SIX5* heterozygote mice (Klesert et al., 2000; Sarkar et al., 2000, 2004; Wakimoto et al., 2002). Moreover, they are consistent with the negative *cis* effect of large CTG expansions on the induction of hypermethylation in transgenic mice (Brouwer et al., 2013; López Castel et al., 2011), pointing to a role for chromatin modification in DM1 pathogenesis. However, it should be noted that hypermethylation patterns in transgenic mice are similar, but not identical, to those observed in DM1 individuals (López Castel et al., 2011) and that, in contrast to our study, the reduced expression in *SIX5* is uncorrelated with expansion size in transgenic mice (Brouwer et al., 2013). The discrepancies between both studies may stem from epigenetic variations that exist between species or may be attributed to transgene integration site effects.

Given the aforementioned relationships, we aimed to mechanistically associate DMR hypermethylation with allele-specific reduction in *SIX5* expression. Initially, we ruled out the possibility that the CTG expansion leads to the reduction in *SIX5* expression by interfering with CTCF binding via hypermethylation, as proposed previously (Filippova et al., 2001). Our findings do not support the proposition that CTCF normally functions as an insulator binding protein at the DM1 locus to guarantee the independent expression of *DMPK* and *SIX5* (Cho et al., 2005). Although this difference may stem from the different cell types employed in each study, our findings are well in line with the report of Brouwer et al. (2013) regarding transgenic mice that showed no effect of CTG expansion size on CTCF binding.

In light of these results regarding CTCF, we questioned whether methylation spreading may exert its effect by an alternative mechanism: by hampering the activity of a putative regulatory element for *SIX5* that resides within the *DMPK* gene. For this purpose, we examined the ability of the DMR to drive reporter gene expression in vivo and in vitro. In vivo, we found that the DMR is capable and sufficient to drive GFP expression in the heads of transgenic fish, predominantly in the epidermis, hindbrain, and branchial arches. We speculate that the partial similarity in the tissue specificity of reporter gene activity in zebrafish to that of *SIX5* expression in the branchial arches of developing mice (10.5 days post-coitum [d.p.c.]) (Heath et al., 1997; Klesert et al., 2000) may possibly explain the lack of facial expression and the difficulties in speech and swallowing that are typical in patients.

In vitro, we demonstrated that the DMR can drive luciferase expression in three different cell lines (293T,COS-7, and HepG2) and markedly enhances reporter gene activity (10- to 15-fold increase) compared with basal activity obtained with either the *SIX5* promoter, the entire region that surrounds the CTG repeats (12 repeats), or the DMR in an opposite orientation. Although our in vitro assays point to the potential role of this regulatory element as a promoter sequence, there is no evidence to support this in publically available datasets in hESCs or any other cell type examined. Therefore, we suggest that this element most likely acts as a proximal enhancer rather than a distal promoter for *SIX5* transcription. Importantly, to establish the relevance of methylation to the regulatory function of this region, we verified that in vitro methylation of the DMR completely silenced reporter gene activity in all examined cell lines. These results indicate that aberrant methylation is, in fact, mechanistically responsible for the *cis* reduction in *SIX5* expression. Lastly, given the frequent involvement of cardiac conduction defects in patients and in *SIX5* heterozygote mice (Martorell et al., 1997; Sarkar et al., 2000; Wakimoto et al., 2002), we illustrate that the DMR is capable of driving eGFP expression in in vitro-derived transgenic cardiomyocytes.

Our results are in line with the identification of this region in publically available data as a regulatory element by FAIRE (formaldehyde-assisted isolation of regulatory elements); DNaseI hypersensitive site mapping; local enrichment in monomethylation of histone 3 at lysine 4 (H3K4me1), acetylation of histone 3 at lysine 27 (H3K27Ac) and Pol2A; and the clustering of tissue-specific transcription factors. In conclusion, our functional assays suggest that *SIX5* activity is finely tuned by the activity of a narrow DMR that is located upstream of the repeats, in the coding region of *DMPK*. This putative regulatory zone may ultimately prove to be an essential target for the treatment of DM1 clinical symptoms that are attributed to *SIX5* downregulation.

Together, this study shows a mechanistic association between CTG expansion size, hypermethylation, and the reduction in *SIX5* expression. It demonstrates how a disease-causing mutation in one gene (CTG expansion in *DMPK*) can affect the expression of a different gene (*SIX5*) by epigenetically modifying a flanking sequence that plays a dual function: a protein coding sequence and a regulatory element for a neighboring gene. Our findings, therefore, clearly illustrate how an unstable repeat expansion leads to the spread of methylation by the loss of the 5′ boundary of a CpG island.

Lastly, this study highlights the power of mutant hESCs in deciphering mechanistic relations that are exclusive in undifferentiated cells, such as the relation between methylation spreading and expansion size. These associations probably would not have been revealed if not for a large

and heterogeneous cohort of mutant hESC lines that were available for analysis. Moreover, this study emphasizes the strength of mutant hESCs in disease modeling by affording the exploration of human-based, disease-relevant tissues, like impaired cardiomyocytes, that are otherwise inaccessible for research.

## EXPERIMENTAL PROCEDURES

### hESC Line Derivation, Characterization, and Spontaneous Differentiation In Vitro

The use of DM1-affected embryos, derived from preimplantation genetic diagnosis (PGD) treatment, for hESC derivation was performed in compliance with protocols approved by the National Ethics Committee. All cell lines were established at the Shaare Zedek Medical Center (87/07), apart from LIS-DM, which was provided by the Racine IVF Unit, Tel-Aviv Sourasky Medical Center (7/04-043). Cell line derivation and characterization were carried out as described previously (Eiges et al., 2007). All DM1 hESC lines were examined for typical characteristics of hESCs (primers for RT-PCR analysis are listed in Table S1).

### Expression of Undifferentiated Specific Markers

Undifferentiated cell cultures were examined for the expression of undifferentiated cell-specific markers by immunostaining using monoclonal mouse OCT4 (Santa Cruz Biotechnology, catalog no. sc-5279, 1:50 dilution) or TRA-1-60 (Santa Cruz, catalog no. sc-21705, 1:50 dilution) together with Cy3-conjugated goat anti-mouse polyclonal antibodies (Jackson Immunostaining, catalog no. 115-035-062, 1:100 dilution). Nuclear staining was performed with Hoechst 33258 (Sigma, catalog no. 861405). Staining for alkaline phosphatase was carried out using an alkaline phosphatase kit (Sigma Diagnostics, catalog no. 86R-1KT), according to the manufacturer's protocol.

### DM1 iPSC Derivation

Four retroviral vectors expressing OCT4, SOX2, KLF4, and c-MYC were packaged individually in 293T cells. Infectious viruses were collected 24 and 48 hr post-transfection and added immediately to primary fibroblasts (catalog no. GM03132A, Coriell Cell Repositories). Four days after infection, the cells were placed on mytomycin C-treated mouse embryonic fibroblasts (MEFs) and maintained in hESC medium. Manual isolation of single clones was carried out approximately 30 days post-transfection, resulting in stable cell lines with a hESC-like morphology.

### RT-PCR

Total RNA was isolated from the cells by TRI reagent extraction, and then 1–2 μg RNA was reverse-transcribed by random hexamer priming and Multi Scribe reverse transcriptase (ABI). Amplification was performed using the primers listed in Table S1 using SuperTherm Taq DNA polymerase (Jain Biologicals). For allele expression ratio analysis, DNA and cDNA PCR products were purified and sequenced on an ABI 3130XL genetic analyzer. Sequencing chromatograms were analyzed using PeakPicker2 software.

### Southern Blot Analysis

Genomic DNAs (10–25 μg) were digested with SacI and HindIII (Fermentas) restriction endonucleases, separated on 0.8% agarose gels, blotted onto Hybond N+ membranes (Amersham), and hybridized with a PCR Dig-labeled pDM576 probe (primers: 5′-GCT AGG AAG CAG CCA ATG AC-3′ and 5′-CAT TCC CGG CTA CAA GGA C-3′). For analysis of DNA methylation, triple digests of SacI and HindIII were performed together with either the methylation-sensitive enzyme HpaII or its methylation-insensitive isoschizomer MspI (Fermentas).

### SNP Analysis

Genomic DNA from each cell line was PCR-amplified (primers are listed in Table S1) and compared with parental DNAs at the SNP site to determine allelic associations with the CTG expansion.

### Bisulfite Sequencing

Genomic DNA (2 μg) was modified by bisulfite treatment (EZ DNA methylation kit, Zymo Research) and amplified by FastStart DNA polymerase (Roche). Amplified products were cloned, and single colonies were analyzed for CpG methylation by direct sequencing (ABI 3130). For pyrosequencing, PCR products were analyzed using PyroMak Q24 (QIAGEN). Primers are listed in Table S1.

### Cardiomyocyte Differentiation

WT and DM1-affected hESC lines were induced to differentiate into human cardiomyocyte cells according to Burridge et al. (2014). Briefly, hESC lines grown on a mouse embryonic fibroblast feeder layer were passaged into Essential 8 medium on vitronectin-coated culture vessels and then passaged with 0.5 mM EDTA. For differentiation, the medium was changed to CDM3, consisting of RPMI 1640 medium (catalog no. 11875, Life Technologies), 500 μg/ml *O. sativa*-derived recombinant human albumin (catalog no. A0237, Sigma-Aldrich), and 213 μg/ml L-ascorbic acid 2-phosphate (Sigma-Aldrich). The medium was changed every other day (48 h). On days 0–2, the medium was supplemented with 6 μM CHIR99021 (LC Laboratories). On day 2, the medium was changed to CDM3 supplemented with 2 μM Wnt-C59 (Selleck Chemicals). The medium was changed to CDM3 on day 4 and every other day. Contracting cells were noted from day 7. The human cardiomyocytes were collected for analysis by TrypLE Express (Life Technologies) and were examined for expression of three cardiac markers (MYH-7, GATA-4, and NKX 2-5). Primers are listed in Table S1.

### Chromatin Immunoprecipitation

Chromatin immunoprecipitation was performed using the Upstate EZ ChIP kit according to the manufacturer's protocol, with slight modifications. In brief, cells were harvested and then fixed, quenched, and washed in 50-ml tubes. Sonication was carried out using Vibra Cell VCX130 with a 3-mm microtip and 30% amplitude in 5 cycles of 10 min and 30 min rest on ice. Immunoprecipitation was performed using an anti-CTCF antibody (Upstate, catalog no. 07-729). Real-time PCR was carried out on an ABI 7900HT instrument (primers are listed in Table S1). ΔΔCt values were normalized according to a positive control (*FXN*) to account for differences in precipitation efficiency.

### Zebrafish Functional Assay

Three fragments (DMR, the *SIX5* promoter, and the entire region) were cloned into the E1b GFP-Tol2 vector containing an E1b minimal promoter, followed by GFP (Li et al., 2010; Birnbaum et al., 2012). Briefly, the zebrafish were injected following standard procedures (Nusslein-Volhard and Dahm 2002) into at least 100 embryos/construct along with Tol2 mRNA (Kawakami, 2005) to facilitate genomic integration. GFP expression was observed and annotated at 48 and 72 hours post-fertilization (hpf). All animal care was approved by the Animal Care and Use Committee of Ben-Gurion University.

### Reporter Gene Assay

Cell transfections of the pGL3-based reporter constructs (primers for vector constructions are listed in Table S1) were performed in three different cell lines (293T, COS-7, and HepG2) in 24-well plates using TransIT-LT1 transfection reagent (MIRUS) according to the manufacturer's recommendations. Cells were cotransfected with 100 ng pGL3-based reporter constructs carrying each of the different fragments or an empty vector and with 10 ng of *Renilla* luciferase plasmid with luciferase driven from a thymidine kinase (TK) promoter. Luciferase activity was measured in triplicate 48 h after transfection using the Dual-Luciferase reporter assay (Promega) in a MiniLumat LB 9506 luminometer (EG&G Berthold). To control for transfection efficiency, firefly luciferase activity was normalized to that of *Renilla* luciferase. p Values were obtained using a two-tailed, non-paired t test ($p = 0.05$).

### In Vitro Methylation

Methylation of the pGL3 vector carrying the DMR was performed by CpG methyltransferase (M.SssI, New England Biolabs). The methylase reaction was performed with the addition of the compound S-adenosylmethionine (SAM). The extent of methyltransferase activity was determined by digestion with the methylation-sensitive endonuclease HpaII and its insensitive isoschizomer MspI.

### hESC Transfection

The DMR region was cloned into a basic eGFP-N1 plasmid, which lacks a promoter or an enhancer sequence. A WT hESC line was transfected with the eGFP-N1-DMR plasmid or an empty plasmid using LT1-TransIT transfection reagent (MIRUS) according to the manufacturer's protocol. Stably transfected hESC clones were generated by the use of 100–150 μg G418. DMR-eGFP-transfected hESC clones were induced to differentiate into functional cardiomyocytes following FACS.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures, one table, and four movies and can be found with this article online at http://dx.doi.org/10.1016/j.stemcr.2015.06.003.

## ACKNOWLEDGMENTS

We thank Dr. Amira Gepstein for assistance with *in vitro* cardiac differentiation, Dr. David Zeevi for critical reading of the manuscript, Dr. Dalit Ben-Yosef for the provision of the LIS-DM-affected hESC line, Prof. Douglas Melton and Prof. Nissim Benvenisty for the provision of WT hESC lines (HES13, HES-B123, and HES-B200), and Dr. Micha Mandel for assistance with the statistical analysis. This work was partly supported by the Israel Science Foundation (to E.R., grant 711/12), a donation from the Abrasba Foundation (to E.R., Gindi family), and a Career Integration Grant (to B.R., grant 630849).

Received: May 31, 2015
Revised: June 15, 2015
Accepted: June 15, 2015
Published: July 16, 2015

## REFERENCES

Aslanidis, C., Jansen, G., Amemiya, C., Shutler, G., Mahadevan, M., Tsilfidis, C., Chen, C., Alleman, J., Wormskamp, N.G., Vooijs, M., et al. (1992). Cloning of the essential myotonic dystrophy region and mapping of the putative defect. Nature *355*, 548–551.

Ben-David, E., Granot-Hershkovitz, E., Monderer-Rothkoff, G., Lerer, E., Levi, S., Yaari, M., Ebstein, R.P., Yirmiya, N., and Shifman, S. (2011). Identification of a functional rare variant in autism using genome-wide screen for monoallelic expression. Hum. Mol. Genet. *20*, 3632–3641.

Birnbaum, R.Y., Clowney, E.J., Agamy, O., Kim, M.J., Zhao, J., Yamanaka, T., Pappalardo, Z., Clarke, S.L., Wenger, A.M., Nguyen, L., et al. (2012). Coding exons function as tissue-specific enhancers of nearby genes. Genome Res. *22*, 1059–1068.

Brook, J.D., McCurrach, M.E., Harley, H.G., Buckler, A.J., Church, D., Aburatani, H., Hunter, K., Stanton, V.P., Thirion, J.P., Hudson, T., et al. (1992). Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. Cell *69*, 385.

Brouwer, J.R., Huguet, A., Nicole, A., Munnich, A., and Gourdon, G. (2013). Transcriptionally Repressive Chromatin Remodelling and CpG Methylation in the Presence of Expanded CTG-Repeats at the DM1 Locus. J. Nucleic Acids *2013*, 567435.

Burridge, P.W., Matsa, E., Shukla, P., Lin, Z.C., Churko, J.M., Ebert, A.D., Lan, F., Diecke, S., Huber, B., Mordwinkin, N.M., et al. (2014). Chemically defined generation of human cardiomyocytes. Nat. Methods *11*, 855–860.

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

De Temmerman, N., Seneca, S., Van Steirteghem, A., Haentjens, P., Van der Elst, J., Liebaers, I., and Sermon, K.D. (2008). CTG repeat instability in a human embryonic stem cell line carrying the myotonic dystrophy type 1 mutation. Mol. Hum. Reprod. *14*, 405–412.

Eiges, R., Urbach, A., Malcov, M., Frumkin, T., Schwartz, T., Amit, A., Yaron, Y., Eden, A., Yanuka, O., Benvenisty, N., and Ben-Yosef, D. (2007). Developmental study of fragile X syndrome using human embryonic stem cells derived from preimplantation genetically diagnosed embryos. Cell Stem Cell *1*, 568–577.

Filippova, G.N. (2008). Genetics and epigenetics of the multifunctional protein CTCF. Curr. Top. Dev. Biol. *80*, 337–360.

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

Heath, S.K., Carne, S., Hoyle, C., Johnson, K.J., and Wells, D.J. (1997). Characterisation of expression of mDMAHP, a homeodomain-encoding gene at the murine DM locus. Hum. Mol. Genet. *6*, 651–657.

Kawakami, K. (2005). Transposon tools and methods in zebrafish. Dev. Dyn. *234*, 244–254.

Klesert, T.R., Otten, A.D., Bird, T.D., and Tapscott, S.J. (1997). Trinucleotide repeat expansion at the myotonic dystrophy locus reduces expression of DMAHP. Nat. Genet. *16*, 402–406.

Klesert, T.R., Cho, D.H., Clark, J.I., Maylie, J., Adelman, J., Snider, L., Yuen, E.C., Soriano, P., and Tapscott, S.J. (2000). Mice deficient in Six5 develop cataracts: implications for myotonic dystrophy. Nat. Genet. *25*, 105–109.

Korade-Mirnics, Z., Tarleton, J., Servidei, S., Casey, R.R., Gennarelli, M., Pegoraro, E., Angelini, C., and Hoffman, E.P. (1999). Myotonic dystrophy: tissue-specific effect of somatic CTG expansions on allele-specific DMAHP/SIX5 expression. Hum. Mol. Genet. *8*, 1017–1023.

Li, Q., Ritter, D., Yang, N., Dong, Z., Li, H., Chuang, J.H., and Guo, S. (2010). A systematic approach to identify functional motifs within vertebrate developmental enhancers. Dev. Biol. *337*, 484–495.

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

Lund, M., Diaz, L.J., Ranthe, M.F., Petri, H., Duno, M., Juncker, I., Eiberg, H., Vissing, J., Bundgaard, H., Wohlfahrt, J., and Melbye, M. (2014). Cardiac involvement in myotonic dystrophy: a nationwide cohort study. Eur. Heart J. *35*, 2158–2164.

Marteyn, A., Maury, Y., Gauthier, M.M., Lecuyer, C., Vernet, R., Denis, J.A., Pietu, G., Peschanski, M., and Martinat, C. (2011). Mutant human embryonic stem cells reveal neurite and synapse formation defects in type 1 myotonic dystrophy. Cell Stem Cell *8*, 434–444.

Martorell, L., Johnson, K., Boucher, C.A., and Baiget, M. (1997). Somatic instability of the myotonic dystrophy (CTG)n repeat during human fetal development. Hum. Mol. Genet. *6*, 877–880.

Nusslein-Volhard, C., and Dahm, R. (2002). Zebrafish (Oxford: Oxford University Press).

Sarkar, P.S., Appukuttan, B., Han, J., Ito, Y., Ai, C., Tsai, W., Chai, Y., Stout, J.T., and Reddy, S. (2000). Heterozygous loss of Six5 in mice is sufficient to cause ocular cataracts. Nat. Genet. *25*, 110–114.

Sarkar, P.S., Paul, S., Han, J., and Reddy, S. (2004). Six5 is required for spermatogenic cell survival and spermiogenesis. Hum. Mol. Genet. *13*, 1421–1431.

Seriola, A., Spits, C., Simard, J.P., Hilven, P., Haentjens, P., Pearson, C.E., and Sermon, K. (2011). Huntington's and myotonic dystrophy hESCs: down-regulated trinucleotide repeat instability and mismatch repair machinery expression upon differentiation. Hum. Mol. Genet. *20*, 176–185.

Sovari, A.A., Bodine, C.K., and Farokhi, F. (2007). Cardiovascular manifestations of myotonic dystrophy-1. Cardiol. Rev. *15*, 191–194.

Spits, C., Seneca, S., Hilven, P., Liebaers, I., and Sermon, K. (2010). Methylation of the CpG sites in the myotonic dystrophy locus does not correlate with CTG expansion size or with the congenital form of the disease. J. Med. Genet. *47*, 700–703.

Steinbach, P., Gläser, D., Vogel, W., Wolf, M., and Schwemmle, S. (1998). The DMPK gene of severely affected myotonic dystrophy patients is hypermethylated proximal to the largely expanded CTG repeat. Am. J. Hum. Genet. *62*, 278–285.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861–872.

Thornton, C.A., Wymer, J.P., Simmons, Z., McClain, C., and Moxley, R.T., 3rd. (1997). Expansion of the myotonic dystrophy CTG repeat reduces expression of the flanking DMAHP gene. Nat. Genet. *16*, 407–409.

Wakimoto, H., Maguire, C.T., Sherwood, M.C., Vargas, M.M., Sarkar, P.S., Han, J., Reddy, S., and Berul, C.I. (2002). Characterization of cardiac conduction system abnormalities in mice with targeted disruption of Six5 gene. J. Interv. Card. Electrophysiol. *7*, 127–135.