Human Molecular Genetics, 2020, Vol. 29, No. 23 3818–3829

doi: 10.1093/hmg/ddaa267
Advance Access Publication Date: 11 January 2021
General Article

GENERAL ARTICLE

Methylated and unmethylated epialleles support variegated epigenetic silencing in Friedreich ataxia

Layne N. Rodden[1,2], Yogesh K. Chutake[1], Kaitlyn Gilliam[1], Christina Lam[1], Elisabetta Soragni[3], Lauren Hauser[4], Matthew Gilliam[5], Graham Wiley[6], Michael P. Anderson[7], Joel M. Gottesfeld[3], David R. Lynch[4] and Sanjay I. Bidichandani[1,2,8,*]

[1]Department of Pediatrics, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA, [2]Oklahoma Center for Neuroscience, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA, [3]Department of Molecular Medicine, The Scripps Research Institute, La Jolla, CA, USA, [4]Division of Neurology, The Children's Hospital of Philadelphia, Philadelphia, PA, USA, [5]Department of Electrical and Computer Engineering, University of Oklahoma, Norman, OK, USA, [6]Arthritis and Clinical Immunology Program, Oklahoma Medical Research Foundation, Oklahoma City, OK, USA, [7]Department of Biostatistics and Epidemiology, Hudson College of Public Health, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA and [8]Department of Biochemistry and Molecular Biology, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA

*To whom correspondence should be addressed at: OU Children's Physician Building, Suite 12100, 1200 Children's Avenue, Oklahoma City, OK 73104, USA. Tel: + 1 4052711358; Fax: +1 4052718697; Email: Sanjay-Bidichandani@ouhsc.edu

Abstract

Friedreich ataxia (FRDA) is typically caused by homozygosity for an expanded GAA triplet-repeat in intron 1 of the *FXN* gene, which results in transcriptional deficiency via epigenetic silencing. Most patients are homozygous for alleles containing > 500 triplets, but a subset (~20%) have at least one expanded allele with < 500 triplets and a distinctly milder phenotype. We show that in FRDA DNA methylation spreads upstream from the expanded repeat, further than previously recognized, and establishes an FRDA-specific region of hypermethylation in intron 1 (~90% in FRDA versus < 10% in non-FRDA) as a novel epigenetic signature. The hypermethylation of this differentially methylated region (FRDA-DMR) was observed in a variety of patient-derived cells; it significantly correlated with *FXN* transcriptional deficiency and age of onset, and it reverted to the non-disease state in isogenically corrected induced pluripotent stem cell (iPSC)-derived neurons. Bisulfite deep sequencing of the FRDA-DMR in peripheral blood mononuclear cells from 73 FRDA patients revealed considerable intra-individual epiallelic variability, including fully methylated, partially methylated, and unmethylated epialleles. Although unmethylated epialleles were rare (median = 0.33%) in typical patients homozygous for long GAA alleles with > 500 triplets, a significantly higher prevalence of unmethylated epialleles (median = 9.8%) was observed in patients with at least one allele containing < 500 triplets, less severe *FXN* deficiency (>20%) and later onset (>15 years). The higher prevalence in mild FRDA of somatic *FXN* epialleles devoid of DNA methylation is consistent with variegated epigenetic silencing mediated by expanded triplet-repeats. The *proportion* of unsilenced somatic *FXN* genes is an unrecognized phenotypic determinant in FRDA and has implications for the deployment of effective therapies.

Received: August 26, 2020. **Revised:** December 1, 2020. **Accepted:** December 9, 2020

Introduction

Friedreich ataxia (FRDA) is the most common recessive ataxia and is clinically characterized by progressive limb and gait ataxia, dysarthria, weakness, fatigue and cardiomyopathy (1). The vast majority of patients (95%) inherit an expanded GAA triplet-repeat sequence in intron 1 of the *FXN* gene from each parent (2). Whereas non-FRDA alleles contain <30 triplets, the expanded GAA sequence in FRDA ranges from 100 to 1500 triplets. The expanded GAA sequence results in *FXN* transcriptional deficiency (5–30% of non-FRDA levels) (3–6), which in turn causes deficiency of frataxin (7), a protein required for efficient Fe-S cluster biogenesis and mitochondrial oxidative phosphorylation (8–10). Most FRDA patients are homozygous for long GAA triplet-repeats that contain >500 triplets (usually, 600–1200 triplets) (2,11–15). Such patients present with a typically progressive FRDA phenotype, with age of symptom onset in the early teens, and premature mortality in their 30s or 40s (11–18). However, a substantial minority (~20%) of patients have at least one expanded allele with <500 triplets. They usually have less severe *FXN* transcriptional deficiency compared with typical FRDA patients (4–6), and a generally milder phenotype, including later onset, slower progression, lower prevalence of cardiomyopathy and longer lifespan (11–20). Indeed, most of the correlation that is observed between the length of the GAA triplet-repeat and either *FXN* transcriptional deficiency or a number of phenotypic measures of disease severity stems from the bimodal distribution of FRDA patients, molecularly distinguished by the presence/absence of at least one allele with <500 triplets. In addition to the milder phenotype, those with relatively short expanded alleles also show remarkable inter-individual variability in *FXN* transcriptional deficiency, age of onset and rate of disease progression (5,6,11,16,19).

Epigenetic silencing in a manner similar to position effect variegation (PEV) was observed in a transgenic mouse model containing a GAA-200 repeat linked to an upstream reporter gene (21). Variegated expression manifested as proportions of cells showing either reporter gene repression or lack of repression, and repression was associated with increased nucleosomal density and reduced promoter accessibility (21). Following this seminal discovery, various signatures of repressive chromatin, including histone trimethylation (H3K27me3, H3K9me3) and hypoacetylation, were identified at the *FXN* locus in FRDA patients (22–33), as was increased nucleosomal density and reduced promoter accessibility (30). Repressive chromatin in FRDA spreads upstream from the expanded GAA triplet-repeat in intron 1, extending up to the *FXN* promoter, and contributes to *FXN* transcriptional silencing via a combination of deficient initiation and elongation (6,27–30,34,35). Although there is incontrovertible evidence for epigenetic silencing of the *FXN* gene in FRDA, there is as yet no evidence of variegated silencing. This is partly due to the inability of conventional chromatin immunoprecipitation assays to achieve single-molecule/single-cell resolution. However, it could also be because the most common expanded alleles in FRDA contain >500 triplets, and variegated silencing was noted with a relatively short GAA repeat (200 triplets) (21). It therefore remains to be determined if the subset of atypically mild FRDA patients who have expanded GAA alleles with <500 triplets have PEV-like silencing, i.e. variable spreading of epigenetic silencing signals that result in a proportion of somatic *FXN* genes being spared.

Increased DNA methylation has also been observed in the immediate vicinity of the expanded GAA triplet-repeat in FRDA (23,24,33,36,37), but its extent and magnitude at the *FXN* locus remain incompletely characterized. The *FXN* promoter, similar to two-thirds of all human promoters, is associated with a CpG island that spans the 5′ untranslated region, exon 1 and the proximal part (5′ end) of intron 1 (Fig. 1A). The high prevalence and density of CpG sites between the expanded GAA repeat and the *FXN* promoter, which is mostly comprised of the CpG island and its shore, make this region a suitable target for spreading of DNA methylation, especially since repressive chromatin forms in this region in FRDA. If DNA methylation were to spread upstream of the expanded GAA repeat in FRDA, the resulting heterochromatin formed in conjunction with repressive histone marks would present a formidable block to transcriptional elongation, as is known to occur in FRDA. DNA methylation could conceivably also impact transcriptional initiation if it were to involve the CpG island and/or *FXN* promoter.

In a large and representative cohort of FRDA patients, we show that DNA methylation spreads upstream from the expanded repeat, considerably further than previously recognized, and establishes an FRDA-specific differentially methylated region (FRDA-DMR), just downstream of the *FXN* CpG island. This indicates that in FRDA a region of heterochromatin is formed upstream of the expanded GAA triplet-repeat in intron 1 consisting of various repressive histone marks plus DNA hypermethylation. It is noteworthy that in FRDA an R-loop is also known to form in this region of intron 1 (38,39). Moreover, bisulfite deep sequencing of the FRDA-DMR permitted detailed characterization of the epiallelic variability among individual somatic *FXN* genes in FRDA. Consistent with a PEV-like silencing mechanism, we found that the prevalence of *FXN* genes that have escaped from DNA methylation (unmethylated epialleles), while rare in typical FRDA patients, was relatively high among those patients with at least one expanded allele containing <500 triplets. The prevalence of these unmethylated epialleles predicted the severity of *FXN* transcriptional deficiency and age of symptom onset, and therefore serves as a discriminatory epigenetic signal that distinguishes the two major phenotypic groups of FRDA patients. These data support a model in which the proportion of *FXN* genes lacking silencing signals in somatic cells plays a role in phenotypic determination, and that the milder and variable phenotype in FRDA patients with at least one short expanded allele likely stems from a mechanism resembling variegated silencing.

Results

DNA methylation spreads upstream from the expanded GAA triplet-repeat to establish the FRDA-DMR

To determine if DNA methylation spreads upstream from the expanded GAA repeat in FRDA, five key regions were selected to cover CpG dinucleotides between the *FXN* promoter and the GAA repeat in intron 1 (Fig. 1A) Amplicon 1 represented the *FXN* TSS/promoter, amplicons 1 and 2 represented the CpG island, amplicons 3 and 4 represented the CpG island shore, and amplicon 5 represented the region adjacent to the Alu element within which the GAA triplet-repeat is located [Fig. 1A; vertical lines denote each of the 98 CpGs spanning this interval (note: the last three CpGs map within the left arm of the Alu element, and #95 is the last one analyzed in amplicon 5); Supplementary Material, Table S1 lists the *FXN* gene coordinates of all five amplicons and the CpGs they contain]. Other groups have previously found evidence of increased methylation in FRDA involving mostly CpGs that map within amplicon 5 (23,24,33,36,37). DNA methylation was analyzed in all five amplicons by bisulfite treatment of genomic DNA isolated from peripheral blood mononuclear cells (PBMCs), lymphoblastoid cell lines (LBCLs) and primary

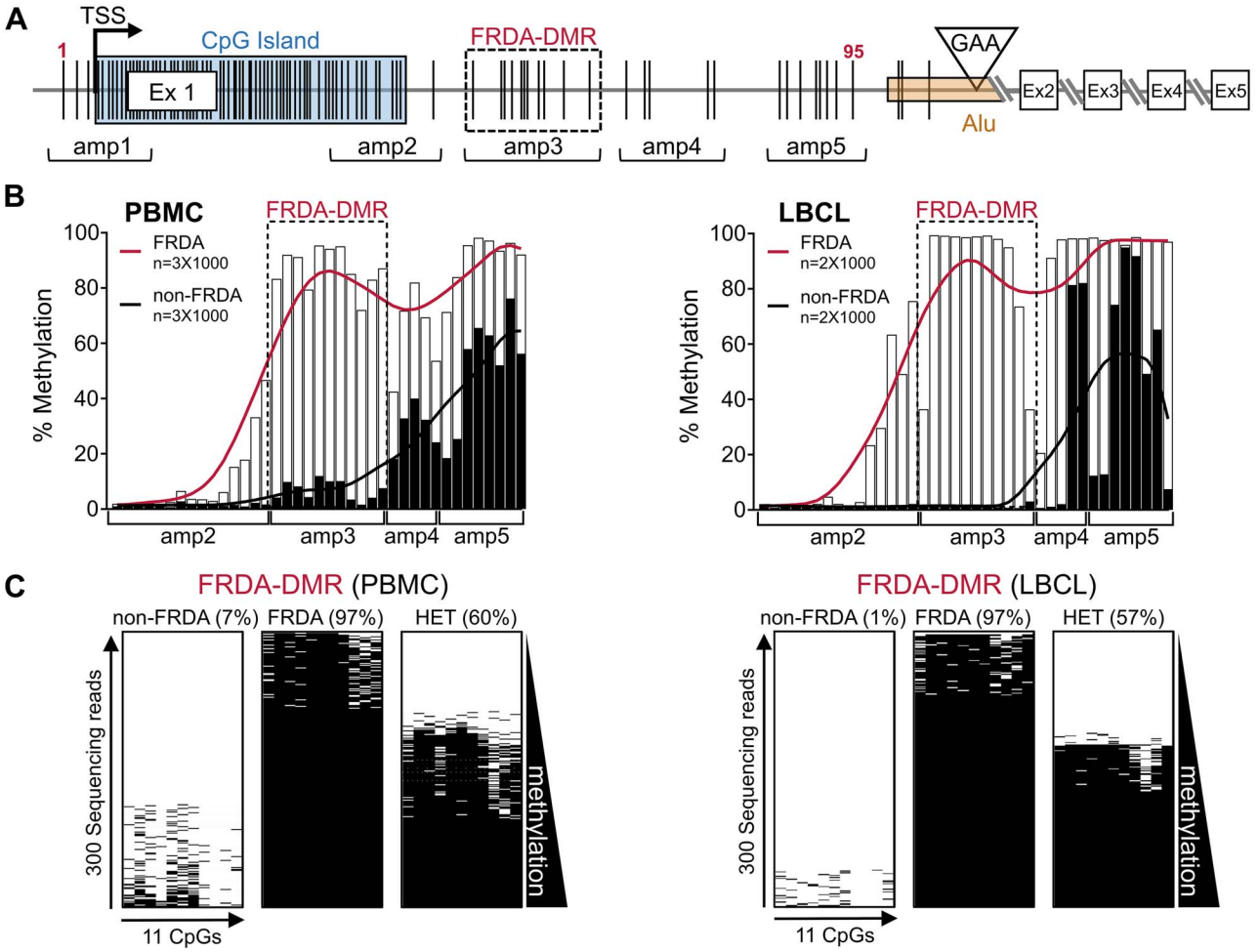

Figure 1. DNA methylation spreads upstream from the expanded GAA triplet-repeat to establish the FRDA-DMR. (**A**) Schematic of the *FXN* gene shows the locations of the five amplicons (amp1-amp5) used for DNA methylation analysis, relative to CpG sites (vertical black lines), transcriptional start site (TSS, at −59), CpG island (light blue box), exon 1 (Ex 1) and the GAA repeat in intron 1 (triangle). The FRDA-DMR, represented by amp3, is indicated by a black dashed box. The numbering of CpG sites starts in the promoter region and continues through to 95 at the end of amp5. The three CpGs downstream of 95, located within the Alu element (orange box), were not analyzed in this study. The exact positions of amps 1–5, and the CpGs contained within them, are listed in Supplementary Material, Table S1. (**B**) DNA methylation (%), calculated from n = 1000 reads at the CpGs in amp2-amp5, is shown for FRDA (white bars; red trendline) and non-FRDA (black bars and trendline) in both PBMCs (n = 3 each) and LBCLs (n = 2 each). Upstream spreading of DNA methylation in FRDA tapers off at the 3′ end of amp2 (3′ end of the CpG island). Amp3 is uniformly hypermethylated in FRDA and almost devoid of methylation in non-FRDA, thus establishing the FRDA-DMR just downstream of the CpG island. (**C**) Bisulfite sequencing reads (n = 300) of the 11 CpGs contained in the FRDA-DMR (amp3), assayed for methylation in *cis*, are stacked vertically and sorted with highest methylation at the bottom (black dash = methylated CpG). Representative results shown for non-FRDA, FRDA and heterozygous carriers (HET) from PBMCs (left) and LBCLs (right) show that the FRDA-DMR is hypermethylated in FRDA (>90% versus < 10% in non-FRDA), and heterozygous carriers have approximately half methylated and half unmethylated *FXN* strands.

fibroblasts from FRDA and non-FRDA controls followed by deep sequencing (*n* = 1000 sequence depth per amplicon; Fig 1B and Supplementary Material, Fig. S1). High levels of DNA methylation were detected in FRDA, compared with non-FRDA controls, primarily in the CpG island shore, represented by amplicons 3 and 4 (Fig. 1B and Supplementary Material, Fig. S1). Although amplicons 4 and 5 showed that at least some CpGs had substantial methylation even in non-FRDA cells, amplicon 3 was uniformly hypermethylated in FRDA and almost devoid of methylation in non-FRDA cells (Fig. 1B and Supplementary Material, Fig. S1). This indicates the establishment of an FRDA-specific differentially methylated region (FRDA-DMR) just downstream of the *FXN* CpG island (dotted boxes in Fig. 1A and B).

Deep sequencing to assess methylation of the 11 CpGs in amplicon 3 in *cis* permitted analysis of the FRDA-DMR in individual *FXN* DNA strands, i.e. somatic *FXN* epialleles. The epiallelic distribution (*n* = 300 stacked strands; black dash = methylation; sorted with highest methylation at the bottom) in FRDA and non-FRDA controls clearly indicates that the FRDA-DMR is hypermethylated in FRDA (>90% vs. < 10% methylation, respectively; Fig. 1C). Additionally, the approximately equal proportion in heterozygous carriers of methylated and unmethylated epialleles indicates that hypermethylation of the FRDA-DMR occurs in *cis* with the expanded GAA repeat. A traditional methylation-sensitive polymerase chain reaction (MS-PCR) assay confirmed that the FRDA-DMR is hypermethylated in FRDA, and a quantitative MS-PCR assay additionally confirmed the ~50% methylation level in heterozygous carriers (Supplementary Material, Fig. S2).

The high level of intronic DNA methylation in FRDA tapers off sharply at the 3′ edge of the CpG island, with the other CpGs in amplicon 2, i.e. further into the CpG island, remaining unmethylated (Fig. 1B and Supplementary Material, Fig. S1). Additionally, all of the CpG sites in amplicon 1 were equally unmethylated in FRDA and non-FRDA cells (Supplementary Material, Fig. S3).

Altogether, these data indicate that the upstream spread of DNA methylation from the expanded GAA repeat in FRDA establishes a hypermethylated FRDA-DMR at the junction of the CpG island and shore, and both the CpG island and *FXN* TSS/promoter remain unmethylated.

The FRDA-DMR is variably hypermethylated in FRDA patients

Polymorphic variability in hypermethylation of the FRDA-DMR was investigated in PBMCs from a consecutive series of 73 FRDA patients (Supplementary Material, Table S2), all of whom were homozygous for expanded GAA repeats (Supplementary Material, Table S3). This cohort size resulted in reasonable ascertainment of the GAA allelic variability typically seen in FRDA (11–15); 22% and 78% had a GAA1 allele (i.e. the shorter of two expanded alleles) of ≤500 and >500 triplets, respectively (Fig. 2A). The median level of FRDA-DMR methylation for the entire cohort was 90%, and a majority of FRDA patients (85%) had methylation levels of >80% (Fig. 2B). In contrast, the level of FRDA-DMR methylation in non-FRDA individuals (n = 10) was <10% (Supplementary Material, Fig. S4; similarly, low levels of methylation were also seen in HEK293T and HeLa cells), and in heterozygous carriers (n = 3) it was ∼65% (Fig. 2B). Although hypermethylation of the FRDA-DMR is the norm in FRDA, 15% of patients had relatively lower levels of methylation, ranging from 56 to 80% (83% = 25th percentile) (Fig. 2B). Stratification by repeat length showed that typical FRDA patients, with both expanded alleles containing >500 triplets, uniformly (98%) had >80% methylation (Fig. 2C). In contrast, a substantial number of those who had a GAA1 allele size of ≤500 triplets (63%) had <80% methylation (Fig. 2D).

Unmethylated epialleles are characteristic of FRDA patients with relatively short expanded alleles

The methylation status of the 11 CpGs within the FRDA-DMR, assayed in *cis*, in 300 individually sequenced *FXN* molecules revealed three types of epialleles: fully methylated (with all 11 CpGs methylated), unmethylated (≤2 CpGs methylated) and partially methylated (with >2 and <11 CpGs methylated) (Fig. 2C and D). Patients with >80% methylation had a high prevalence of fully methylated epialleles (and low levels of unmethylated epialleles), and those with <80% methylation had a higher prevalence of unmethylated epialleles (and relatively low levels of fully methylated epialleles) (Fig. 2C–E). Indeed, when methylation was below 80%, epiallelic variability was mostly driven by the prevalence of unmethylated epialleles, and variability above 80% was mostly driven by the prevalence of fully methylated epialleles (Fig. 2E). Partially methylated epialleles did not show any particular pattern, other than a decrease in prevalence at very high methylation levels, indicating that increase in methylation above 90% occurs via replacement of partially methylated epialleles by fully methylated epialleles (unmethylated epialleles were essentially nonexistent above 90% methylation) (Fig. 2E).

As expected, the level of methylation in the FRDA-DMR correlates well with GAA repeat length (Fig. 2F). Most of the variability in total methylation is seen when the combined size of both expanded alleles (GAA1 + GAA2) is <1500 triplets, and this is mostly due to patients with a GAA1 allele of ≤500 triplets (Fig. 2F). However, the distinguishing feature of FRDA with a GAA1 allele of ≤500 or >500 triplets was the relative prevalence of unmethylated epialleles (median = 9.8% versus 0.33%, respectively; Mann–Whitney test $P < 0.0001$; Fig. 2G). It was rare for someone with a GAA1 allele of >500 to have >5% unmethylated epialleles (95% had <5%; Fisher's exact text, $P < 0.0001$). In contrast, those with a GAA1 allele of ≤500 triplets had a significantly higher chance of having ≥5% unmethylated epialleles (81% of such patients; Fisher's exact text, $P < 0.0001$) (Fig. 2G).

Therefore, in contrast with the relative homogeneity of high FRDA-DMR methylation (∼90%) and low prevalence of unmethylated epialleles (0–5%) in individuals with both expanded alleles containing >500 triplets, FRDA patients with a GAA1 allele of ≤500 triplets have lower overall methylation (<80%) and a higher prevalence of unmethylated epialleles (typically >5%; up to 32%), albeit with considerable inter-individual variability. The higher prevalence of unmethylated *FXN* epialleles in the FRDA-DMR is a novel epigenetic feature of FRDA patients with at least one expanded allele containing ≤500 triplets.

Role of FXN methylation and unmethylated epialleles in transcriptional deficiency in FRDA

FXN transcript levels were measured in the initial 50 (of 73) FRDA patients recruited for this study (Fig. 3A). The median *FXN* transcript level in this cohort, relative to heterozygous carriers (n = 3; set at 50%), was 15%, with most patients expressing <20% (22% = 75th percentile). FRDA-DMR methylation in this subset of patients (Fig. 3B), with a median value of 89% (range: 56–97%), was comparable to the larger FRDA cohort of 73 (Fig. 2A; median = 90%; range: 56–97%; Mann–Whitney test P=0.53), indicating that it is a representative subpopulation. Methylation levels were measured at all CpG sites spanning amplicons 2–5, and these were individually assessed for correlation with *FXN* transcript levels in FRDA. This unbiased approach revealed that *FXN* transcript levels correlated specifically with all of the CpG sites within the FRDA-DMR (CpG #72–82; Fig. 3C), and this correlation weakened and eventually disappeared at CpGs outside the FRDA-DMR, thus localizing the correlation within the FRDA-DMR. A CpG site in the vicinity of the expanded GAA repeat that was previously shown to correlate with *FXN* deficiency (37) was also detected in our cohort (#90 in Fig. 3C).

FXN transcript level in FRDA was inversely correlated with methylation in the *FXN*-DMR (Fig. 3D). Those with both expanded alleles of >500 triplets formed a homogeneous group with high levels of methylation (>80%) and <20% *FXN* transcript. In fact, most of the correlation between methylation and *FXN* transcript was effectively driven by the subset of FRDA patients with a GAA1 allele of ≤500 triplets (Fig. 3D, Supplementary Material, Table S4). *FXN* transcript levels also correlated well with the prevalence of unmethylated epialleles in the FRDA-DMR (Fig. 3E). However, again, this correlation was mostly driven by the subset of patients with GAA1 alleles of ≤500 triplets, who had a much higher prevalence of unmethylated epialleles (>5%) and higher *FXN* transcript levels (>20%) (Fig. 3E, Supplementary Material, Table S4). In contrast, fully methylated epialleles did not show any appreciable correlation with *FXN* transcript levels in patients with a GAA1 allele size of ≤500 or >500 triplets (Supplementary Material, Table S4). These data indicate that the correlation between GAA repeat length and *FXN* transcriptional deficiency is largely a function of those patients with a GAA1 allele of ≤500 triplets and can be explained, at least in part (R^2 = 0.5–0.6), by the level of FRDA-DMR methylation, and specifically the prevalence of unmethylated epialleles in the FRDA-DMR.

In FRDA, the repeat length is known to correlate with *FXN* transcript levels (4–6), and because it also correlates with FRDA-DMR methylation, multiple regression [least absolute

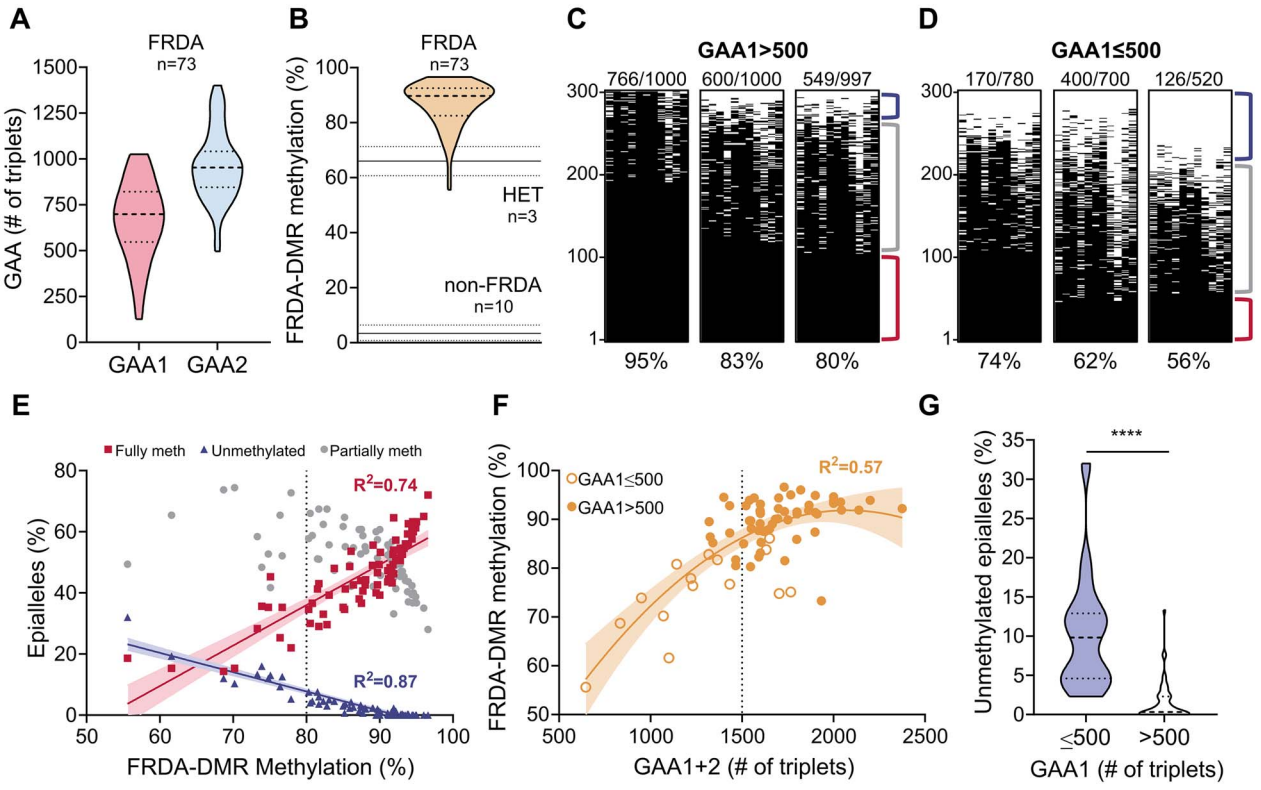

Figure 2. The FRDA-DMR is variably hypermethylated in FRDA, and patients with shorter expanded repeats have a higher prevalence of unmethylated *FXN* epialleles. (**A**) Size distribution of triplet-repeats (GAA1 and GAA2 represent the shorter and longer allele, respectively) and (**B**) the range of methylation in the FRDA-DMR in our cohort of 73 FRDA patients homozygous for the expanded GAA repeat (the heavy dotted line and the two light dotted lines in the violin plots represent the median and 25th/75th percentiles, respectively). Average methylation values for three heterozygous carriers (HET) and three non-FRDA controls (and standard deviation) are indicated by horizontal lines. (**C**) Hypermethylation and epiallelic variability of the FRDA-DMR in patients with a GAA1 of > 500 triplets showing a higher proportion of fully methylated epialleles (red bracket) and few unmethylated epialleles (blue bracket). (**D**) Hypermethylation and epiallelic variability of the FRDA-DMR in FRDA patients with a GAA1 of ≤ 500 triplets showing a higher proportion of unmethylated epialleles (blue bracket) and fewer fully methylated epialleles (red bracket). Both groups of patients show similar levels of partially methylated epialleles (gray bracket). (**E**) When FRDA-DMR methylation is relatively low (<80%), the epiallelic variability is mostly determined by the prevalence of unmethylated epialleles, and with high levels of methylation (>80%) this is mostly driven by the prevalence of fully methylated epialleles. R^2 values are from Pearson correlation and in both cases $P < 0.0001$ (95% confidence intervals are indicated). (**F**) The correlation of FRDA-DMR methylation with total repeat length (GAA1 + GAA2) indicates that most of the correlation is due to patients with a GAA1 of ≤ 500 triplets. R^2 value is from second order polynomial correlation, and the 95% confidence interval is indicated. (**G**) Patients with a GAA1 of ≤ 500 triplets have a significantly higher prevalence of unmethylated epialleles compared with patients with a GAA1 of > 500 triplets (median = 9.8% vs. 0.33%; Mann–Whitney test $P < 0.0001$).

shrinkage and selection operator (LASSO)] was performed to determine the relative contribution of *FXN* DNA methylation on *FXN* transcriptional deficiency. In addition to the length of the GAA repeats (GAA1, GAA2, GAA1 + GAA2), *FXN* methylation was analyzed both within the FRDA-DMR and immediately upstream, as it encroaches into the *FXN* CpG island (M1–5 in Fig. 3C). This revealed that methylation of the 5′ portion of the FRDA-DMR (M4) explained half of the variability in *FXN* transcript levels in FRDA, i.e. quantitatively similar to the predictive value of the expanded GAA repeats (Fig. 3F). This is of particular interest because GAA repeat length is presently the only known quantitative correlate of transcriptional deficiency in FRDA. This indicates that FRDA-DMR methylation provides additional predictive value over GAA repeat length alone, and perhaps contributes to *FXN* transcriptional deficiency by ways that are both dependent on and independent of repeat length.

Role of FXN methylation and unmethylated epialleles in determining age of onset in FRDA

Age of onset was recorded for 71 of the 73 study participants. The median age of onset was 11 years, and 22% of the patients in our cohort had an age of onset ≥15 years (Fig. 4A). The analysis of methylation at all individual CpG sites spanning amplicons 2–5 showed that the age of onset correlated specifically with all of the CpG sites in the FRDA-DMR (CpG #72-82; Fig. 4B), in a manner that mimicked the correlation with transcript levels (Fig. 3C). CpGs in the vicinity of the expanded GAA repeat that were previously found to correlate with age of onset in FRDA (36,37) were also detected here (CpG #89, #90 and #94; Fig. 4B). Age of onset in FRDA was inversely correlated with total methylation in the *FXN*-DMR (Fig. 4C), which is mostly driven by the prevalence of unmethylated epialleles in the FRDA-DMR (Fig. 4D, Supplementary Material, Table S5). In contrast, fully methylated epialleles did not show any appreciable correlation with age of onset in patients with a GAA1 allele size of ≤ 500 or > 500 triplets (Supplementary Material, Table S5). Those with both expanded alleles containing > 500 triplets formed a homogeneous group with earlier age of onset <15 years, high levels of methylation (>80%) and low prevalence of unmethylated epialleles (0–5%) (Fig. 4C and D). Conversely, those with a GAA1 allele of ≤ 500 triplets were more variable, showed later ages of onset (>15 years), lower levels of methylation (<80%) and higher prevalence of unmethylated epialleles (>5%) (Fig. 4C and D). Indeed, it was

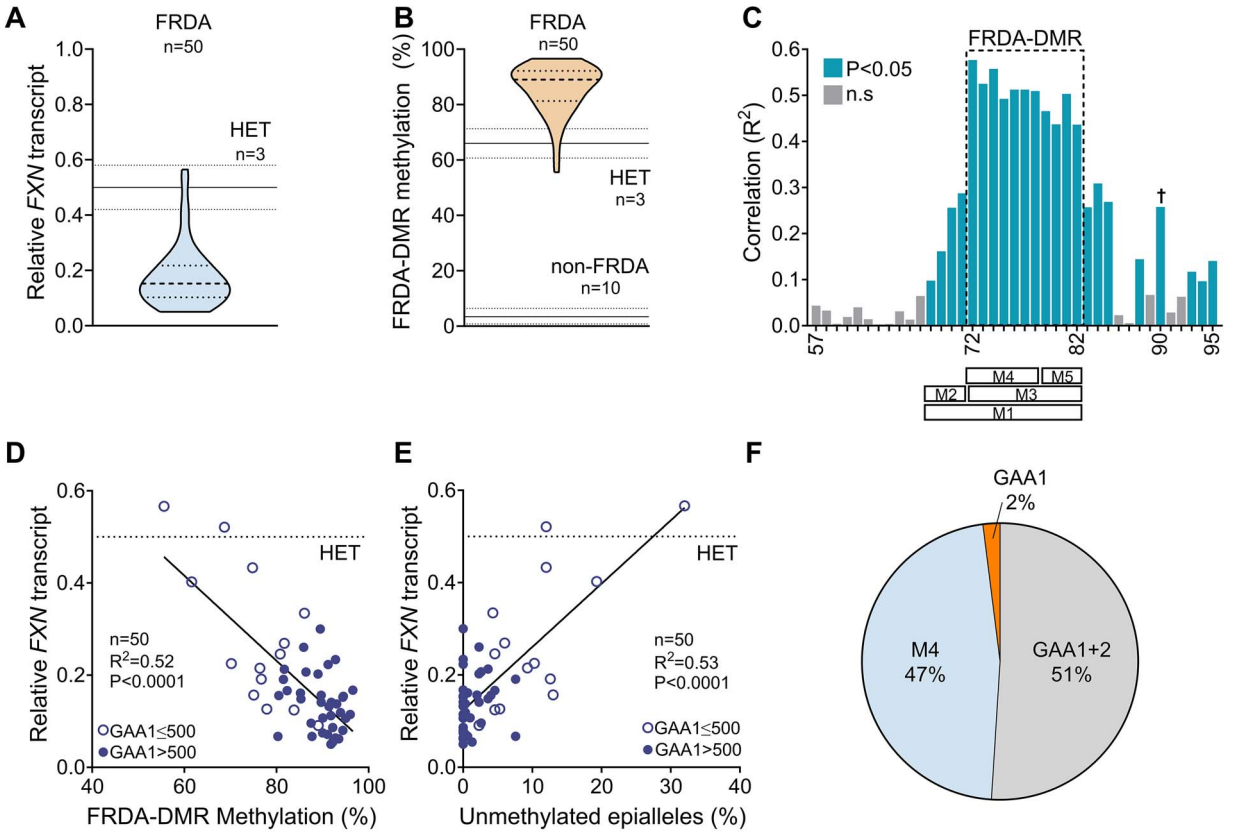

Figure 3. FRDA-DMR methylation is predictive of *FXN* transcriptional deficiency in FRDA. (**A**) The range of *FXN* transcript levels in 50 FRDA patients, relative to three heterozygous carriers (HET; set to 0.5). (**B**) The range of FRDA-DMR methylation level in the cohort of 50 patients, relative to three heterozygous carriers (HET) and three non-FRDA controls (the heavy dotted line, and the two light dotted lines in the violin plots represent the median and 25th/75th percentiles, respectively). (**C**) Bar graph showing R^2 values (Pearson) for the correlation of residual *FXN* transcript levels in the 50 patients with methylation at each CpG site in amps 2–5. Note the clustering of CpG sites showing significant correlation (blue bars), with R^2 values ~0.5, in the FRDA-DMR (dashed box). †Evans-Galea et al. (37) reported significant correlation of *FXN* transcript deficiency with methylation at CpG #90. (**D**) *FXN* transcript deficiency in FRDA is inversely correlated with FRDA-DMR methylation and (**E**) directly correlated with the prevalence of unmethylated epialleles, and in both cases this correlation is mostly driven by patients with a GAA1 of ≤ 500 triplets. (**F**) LASSO regression analysis determined that FRDA-DMR methylation and repeat lengths contribute equally to *FXN* deficiency in FRDA. Multiple regression considered GAA1, GAA2, GAA1 + GAA2 and DNA methylation in and around the FRDA-DMR (see boxes indicating M1–5 in **C**; the CpGs included were M1 = 68–82, M2 = 68–71, M3 = 72–82, M4 = 72–78, M5 = 79–82). Methylation in M4 and the total repeat length (GAA1 + GAA2) were the strongest drivers of *FXN* transcript deficiency.

highly unlikely for someone with < 80% methylation or > 5% unmethylated epialleles to have an age of onset ≤10 years, which otherwise comprised 44% of our cohort. These data indicate that age of onset is explained, at least in part ($R^2 \sim 0.5$), by the level of FRDA-DMR methylation and the prevalence of unmethylated epialleles in the FRDA-DMR.

Expanded repeat-dependent hypermethylation of the FRDA-DMR in FRDA neurons

Differentiated neurons from iPSCs of three FRDA patients (homozygous for the expanded GAA repeat; Supplementary Material, Fig. S5) and three non-FRDA controls (40) were tested for hypermethylation of the FRDA-DMR. FRDA neurons showed very high levels of methylation in the FRDA-DMR (>90%), absence of unmethylated epialleles and a substantial number of fully methylated epialleles (44–63%) (Fig. 5A). Non-FRDA controls, in contrast, had much lower levels of methylation (19–32%), absence of fully methylated epialleles and a substantial number of unmethylated epialleles (34–66%) (Fig. 5B). Neurons from a seamlessly corrected, isogenic iPSC control (40) showed a methylation pattern that was identical to the non-FRDA state (Fig. 5C; FA3* is homozygous for GAA-6, and otherwise genetically identical to FA3, which is homozygous for expanded repeats with 510 and 720 triplets; Supplementary Material, Fig. S5). This indicates that hypermethylation of the FRDA-DMR is dependent on the expanded GAA repeat. Neurons from non-FRDA controls were noted to have a higher proportion of partially methylated epialleles compared with other cell types (PBMCs, lymphoblastoid cells and fibroblasts) from non-FRDA controls. However, partially methylated epialleles were equally prevalent in FRDA and non-FRDA neurons, and the main distinguishing feature of the two phenotypic states was the reciprocal distribution of fully methylated and unmethylated epialleles (Fig. 5D).

Discussion

FXN DNA methylation in FRDA has been evaluated prior to this study (23,24,33,36,37). However, in previous studies methylation analysis was limited to the region flanking the expanded GAA repeat, thus largely focusing on the CpGs represented in amplicon 5 in our study. Consequently, FRDA-specific hypermethylation of the FRDA-DMR has remained undetected. The

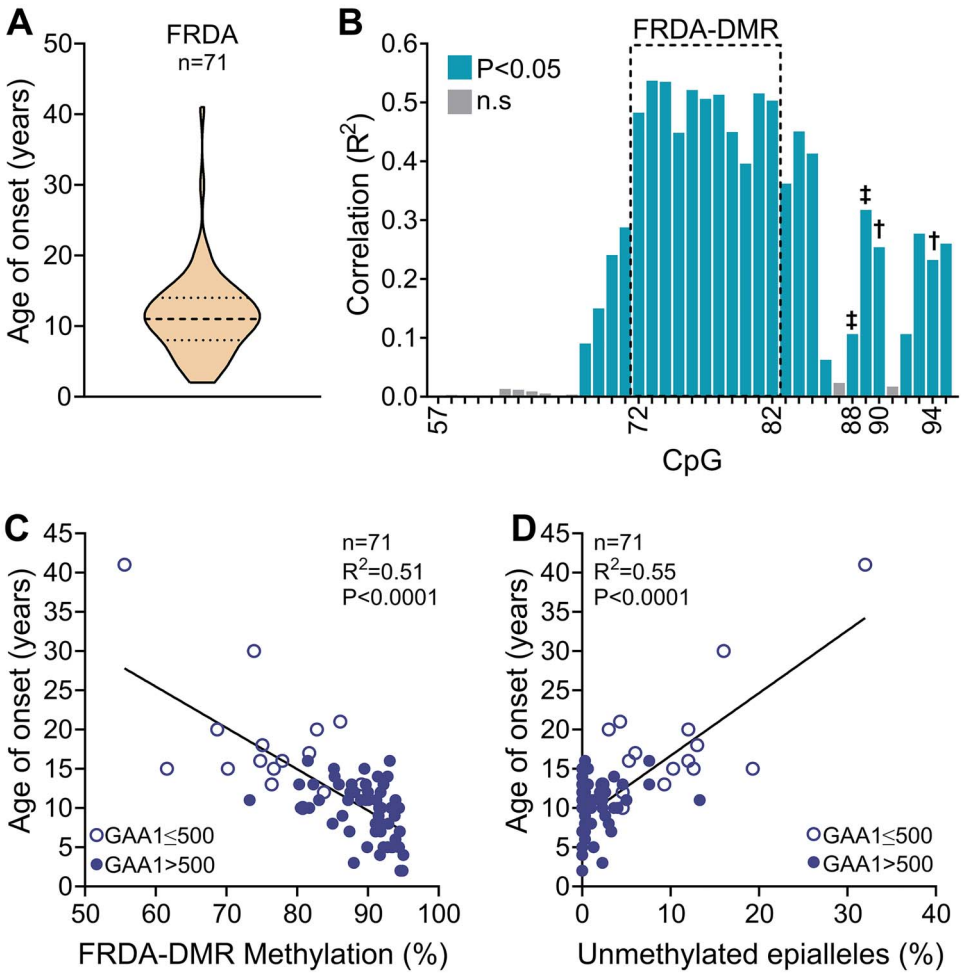

Figure 4. FRDA-DMR methylation is predictive of age of onset in FRDA. (**A**) The range of age of onset is shown for 71 FRDA patients (the heavy dotted line, and the two light dotted lines in the violin plot represent the median and 25th/75th percentiles, respectively). (**B**) Bar graph showing R^2 values (Pearson) for the correlation of age of onset in the 71 patients with methylation at each CpG site in amps 2–5. Note the clustering of CpG sites showing significant correlation (blue bars), with R^2 values ~0.5, in the FRDA-DMR (dashed box). †Evans-Galea et al. (37) reported significant correlation of age of onset and methylation at CpGs 90 and 94. ‡Castaldo et al. (36) reported significant correlation of age of onset and methylation CpGs 88 and 89. (**C**) Age of onset in FRDA is inversely correlated with FRDA-DMR methylation and **(D)** directly correlated with the prevalence of unmethylated epialleles, and in both cases this correlation is mostly driven by patients with a GAA1 of ≤500 triplets. R^2 values are from Pearson correlations.

present study of 73 individuals, with a representative size range of expanded GAA alleles, is the first comprehensive analysis of *FXN* DNA methylation spanning key portions of the *FXN* gene implicated in gene silencing in FRDA. Deep sequencing of the FRDA-DMR as a single amplicon further increased the informativeness of our dataset by allowing the visualization of methylation patterns in a large number of individual *FXN* DNA strands. These may be considered surrogates for individual *FXN* genes, and therefore reflective of the distribution of *FXN* epialleles in somatic cells. The obvious caveat of this assumption is that the same 'gene' may be represented more than once as a result of *in vitro* reamplification, although this pitfall may have been somewhat mitigated by the analysis of a large number of individual strands afforded by the deep sequencing assay. Altogether, our dataset allowed analysis of the extent of DNA methylation in the proximal part of the *FXN* gene and provided a glimpse into the distribution of epiallelic variability in individual somatic *FXN* genes/cells, in a large cohort that accounted for the natural diversity of expanded GAA alleles in FRDA.

The data indicate that DNA methylation spreads upstream from the expanded GAA repeat in FRDA and forms a distinct region of DNA hypermethylation at the junction of the *FXN* CpG island and its shore. How this hypermethylated FRDA-DMR contributes to *FXN* gene silencing is presently unclear. However, the discovery that DNA hypermethylation overlaps with markers of repressive chromatin in the CpG island shore (22–30) provides further evidence of heterochromatin formation in FRDA. This heterochromatin plus an R-loop structure also known to form in the CpG island shore (38) are likely to pose a formidable block to transcriptional elongation, which is known to occur in FRDA (6,27–30,34,35). Additionally, CpG island shore methylation plays a role in regulating transcription (41–43), and there is some evidence that the *FXN* CpG island shore contains regulatory elements (44). Therefore, the hypermethylated FRDA-DMR could possibly interfere with *FXN* transcriptional regulation. It is noteworthy that with almost no CpG to TpG conversion throughout much of primate evolution (Supplementary Material, Fig. S6), the FRDA-DMR has essentially remained unmethylated over tens of millions of years, thus highlighting the anomaly of the hypermethylated state seen in FRDA. However, because the *FXN* TSS/promoter and CpG island are largely spared from DNA methylation (except for limited FRDA-specific encroachment at the 3′ end of

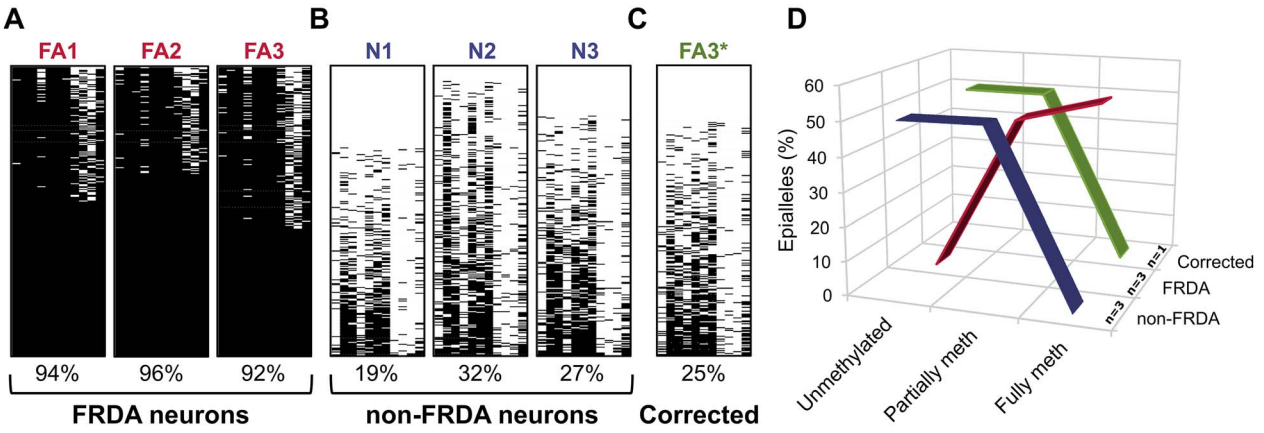

Figure 5. FRDA-specific and expansion-dependent hypermethylation of the FRDA-DMR in neurons. (**A**) The FRDA-DMR is hypermethylated in iPSC-derived neurons from three FRDA patients (FA1, FA2, FA3), as compared with (**B**) three non-FRDA controls (N1, N2, N3). The expanded GAA alleles in the three patient-derived neurons are as follows: FA1 = 550/720; FA2 = 620/620; FA3 = 510/720. (**C**) The methylation pattern in neurons from a seamlessly corrected, isogenic iPSC control (FA3*) reverted to the non-FRDA state. FA3* is homozygous for GAA-6, and otherwise genetically identical to FA3. (**D**) FRDA neurons have no unmethylated epialleles and a high prevalence of fully methylated epialleles. Neurons from non-FRDA controls and from the isogenically corrected control showed the reciprocal pattern of absence of fully methylated epialleles and a higher prevalence of unmethylated epialleles.

the CpG island; Supplementary Material, Fig. S7), it is uncertain if DNA hypermethylation contributes directly to the known deficiency of *FXN* transcriptional initiation in FRDA (5,28,30,34,45). Unfortunately, the hypermethylated FRDA-DMR was resistant to our attempts to induce demethylation via the treatment of patient-derived cells in culture with 5-azacytidine and 5-aza-2′-deoxycytidine (decitabine/DAC), and also via knockdown of DNA methyltransferases.

Our data also show that the majority of FRDA patients, those with > 500 triplets in both expanded alleles, are relatively homogeneous in terms of having high levels of methylation (~90%) and a low prevalence of unmethylated epialleles (0–5%) in the FRDA-DMR. However, unmethylated epialleles seem to be a defining feature of the subset of FRDA patients with GAA1 alleles containing ≤ 500 triplets. We previously showed by small pool PCR analysis of thousands of individual *FXN* molecules that somatic instability in PBMCs results in reversion to non-FRDA allele lengths (<60 triplets) in 0.29% cells, and very large contractions (to < 167 triplets) in 2.3% cells (46). This indicates that homozygous reversion would be seen in $< 10^{-5}$ somatic cells, and homozygosity for large contractions would be seen in $< 10^{-3}$ somatic cells. This level of reversion is too low to account for the ~10% unmethylated epialleles we see in FRDA patients with a GAA1 allele containing ≤ 500 triplets. Furthermore, expanded alleles with ≤ 500 triplets typically undergo small length changes compared with GAA1 alleles with > 500 triplets, which show a relatively strong bias for large contractions (46). However, methylation levels show an opposite pattern, i.e. high prevalence of unmethylated epialleles in FRDA patients with a GAA1 of ≤ 500 triplets and very low prevalence of unmethylated epialleles when the GAA1 allele contains > 500 triplets. It is therefore unlikely that the higher prevalence of unmethylated epialleles observed in FRDA patients with GAA1 alleles containing ≤ 500 triplets is due to somatic instability of the expanded GAA triplet-repeat.

So far, the length of the GAA repeat, especially the GAA1 allele with ≤ 500 triplets, is the strongest known predictor of the age of onset, accounting for ~50% of the variability (11–15,19). This is likely because of the repeat length-dependent transcriptional deficiency seen in FRDA, which is also mostly determined by GAA1 alleles containing ≤ 500 triplets (4–6). Our data indicate that it is the prevalence of unmethylated epialleles that drives the predictive value of FRDA-DMR methylation in determining *FXN* transcript levels and age of onset in FRDA, explaining ~50% of the variability. An implication of this observation is that unmethylated epialleles do not contribute to *FXN* transcriptional deficiency in the way that fully methylated epialleles do. This is suggestive of a PEV-like silencing model, i.e. somatic mosaicism for functionally silenced and non-silenced epialleles (21,47). That unmethylated epialleles are more prevalent in people with a GAA1 allele of ≤ 500 triplets is also consistent with the observation of PEV-like silencing in a transgenic mouse model with a GAA-200 allele (21). Our data therefore support the existence of PEV-like silencing in FRDA, with its clinical impact being the milder phenotype in patients with relatively short expanded alleles (≤500 triplets) and make a case for single-cell analysis to search for direct evidence of variegated silencing.

A simple model of PEV-like silencing in FRDA is presented in Figure 6. Ten cells per hypothetical person are shown, with each semicircle representing a somatic *FXN* epiallele, which is either methylated (shaded) or unmethylated. A non-FRDA control, heterozygous carrier and a typical FRDA patient with both expanded alleles containing > 500 triplets show homogeneous distributions of *FXN* epialleles, based on the normal allele (GAA-8) remaining unmethylated and the expanded allele (GAA-800) becoming fully and uniformly methylated (Fig. 6A). However, in patients with at least one expanded GAA allele containing ≤ 500 triplets, a variable proportion of the cells contains unmethylated *FXN* epialleles (Fig. 6B). In this model, cells with at least one unmethylated *FXN* epiallele would be expected to function normally as in the case of all the cells in the heterozygous carrier. If this model is even partially correct, then we have to contend with a new explanation for the milder FRDA phenotype seen in patients with shorter expanded alleles. So far, the assumption has been that such patients uniformly express higher residual *FXN* transcript levels per cell, but this model predicts that the proportion of cells spared from epigenetic silencing may also contribute to the FRDA phenotype. There are also implications for the effective deployment of therapies designed to reactivate the silenced *FXN* genes (22,33,35,39,48), and for protein/gene

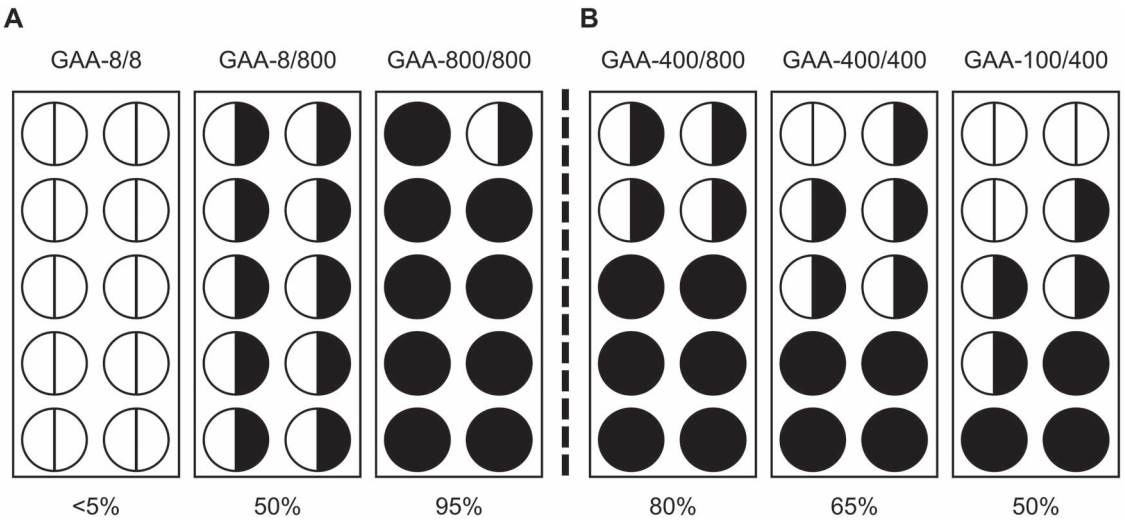

Figure 6. Model of variegated silencing in FRDA. Ten cells per hypothetical person are shown (circles), with each semicircle representing a somatic *FXN* epiallele (methylated = shaded). (**A**) A non-FRDA control (GAA-8/8), heterozygous carrier (GAA-8/800), and a typical FRDA patient with both expanded alleles containing > 500 triplets (GAA-800/800) show homogeneous distributions of *FXN* epialleles, based on the normal allele (GAA-8) remaining unmethylated and the expanded allele (GAA-800) becoming fully methylated. This is consistent with FRDA-DMR methylation of < 10% in non-FRDA, ~ 50% in heterozygotes and > 90% in typical FRDA patients, who also have very few, if any, unmethylated epialleles. (**B**) FRDA patients with at least one allele containing ≤ 500 triplets have FRDA-DMR methylation levels of 50–80% and show a variable proportion of cells with unmethylated *FXN* epialleles. This model predicts that the proportion of cells spared from epigenetic silencing, i.e. variegated silencing, is a determinant of the milder FRDA phenotype seen in FRDA patients with relatively short expanded repeats.

replacement therapies (49,50) in FRDA. The obvious goal of these therapies is to raise the level of frataxin protein to at least the levels seen in heterozygous carriers in all relevant cell types. This model, if true, indicates that the correction of frataxin levels in a proportion of cells may also be expected to yield measurable phenotypic benefit.

Materials and Methods

Study participants and human samples

Peripheral blood samples were collected in purple-top/EDTA tubes from 73 study subjects (with a confirmed DNA diagnosis of FRDA with homozygous expanded GAA triplet-repeat alleles; Supplementary Material, Table S2), 3 heterozygous carriers and 10 unaffected individuals, at the Children's Hospital of Philadelphia (CHOP). Blood samples were shipped overnight on ice via courier service for processing and analysis at the University of Oklahoma Health Sciences Center (OUHSC) in Oklahoma City. Research protocols were approved by the institutional review boards at both institutions (CHOP IRB# 01-002609 and OUHSC IRB# 8071). LBCLs (FRDA: GM16197, GM16798, GM16210; non-FRDA controls: GM22647, GM22671; heterozygous carrier: GM15849) and primary fibroblasts (FRDA: GM03665; non-FRDA: GM07492), were obtained from Coriell Institute for medical research. iPSC-derived neurons (FRDA and non-FRDA) and isogenically corrected iPSC neurons have been described previously (40).

Isolation of DNA and RNA

PBMCs were isolated using Ficoll-Paque PLUS (GE Healthcare). The 'buffy coat' containing PBMCs was carefully collected, washed with phosphate buffered saline (PBS), incubated with erythrocyte lysis buffer (Buffer EL, Qiagen), washed with PBS, and incubated in lymphocyte growth medium (RPMI with Glutamax, 15% FBS, pen/strep) at 37°C in 5% CO_2 for 24–48 h. The average yield was 0.72 million cells/ml of whole blood. Genomic DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen). Total RNA was extracted using the RNeasy Mini Kit (Qiagen).

Measurement of relative FXN transcript level

Total RNA (0.4 µg) was reverse transcribed using the QuantiTect® reverse transcription kit (Qiagen). Transcript levels were quantified by real-time PCR with primers spanning the splice junction of *FXN* exons 3 and 4, relative to expression of *RPL27* using the ΔΔCt method. Power SYBR green PCR Master Mix (Applied Biosystems) was used on a Roche LightCycler® 96 System. Primer sequences and reaction conditions were as previously described (22,51).

Analysis of DNA methylation by bisulfite deep sequencing

Genomic DNA (0.5 µg) was sonicated using the EpiShear Probe Sonicator (Active Motif), and bisulfite converted using either the Active Motif bisulfite conversion kit or the Zymo EZ DNA Methylation Kit, following the respective manufacturer's instructions. Briefly, DNA was incubated with bisulfite reagent for 12 h at 50°C, desulphonated and column purified. Five sets of primers were designed, for each of five amplicons, to capture the CpG dinucleotides between the *FXN* TSS/promoter and the GAA triplet-repeat (see Fig. 1A; Supplementary Material, Table S1 lists primer sequences, positions in the *FXN* gene, and the CpGs assayed in each amplicon). Amplicons were designed to amplify the antisense strand, post-bisulfite conversion. The forward primers in each amplicon contain four 'Ns' (N = A/T/C/G) to increase library diversity (Supplementary Material, Table S1). PCR was performed with AmpliTaq Gold DNA Polymerase with Buffer II (amplicons 1–4 = 2 mM $MgCl_2$; amplicon 5 = 3 mM $MgCl_2$; Applied Biosystems) for 40 cycles using annealing temperatures ranging from 50 to 56°C. Amplicons were column purified and pooled in equimolar amounts for library preparation. Dual indexed

libraries were generated (Nextera XT Index Kit v2 Set A, Illumina) using low-cycle PCR (8 cycles) to incorporate the index sequences (KAPA HiFi HotStart PCR Kit, Kapa Biosystems). Each pool of indexed amplicons was purified (AMPure XP; Beckman-Coulter) and quantified by the Qubit dsDNA Broad Range Assay Kit (Invitrogen/Life Technologies), and further pooled in equimolar amounts to make the final library. The quality and quantity of the final library was analyzed using the Kapa Library Quantification Kit (Kapa Biosystems) using real-time PCR. The library was denatured in 0.1 N NaOH and diluted in HT1 hybridization buffer (Illumina) to 1.4 pM. Due to the low nucleotide diversity of bisulfite DNA, the library was spiked with denatured and diluted (25%) PhiX genomic DNA (Illumina) to improve the efficiency of base calling. The MiniSeq Mid Output 300-cycle reagent cartridge (Illumina) was used for paired-end sequencing, and FASTQ files were analyzed for DNA methylation.

Bioinformatics and display of methylation data

Raw FASTQ reads were trimmed (Trimmomatic), and the forward and reverse reads were merged (bbmerge.sh), discarding inappropriate read lengths. The reads for each sample were aligned to the *in silico* bisulfited reference amplicon sequences (Bowtie2), and parsed using Samtools from the aligned BAM file. A random subset of reads was selected from each BAM file. Samtools mpileup was used to generate a VCF file for each amplicon of each respective sample. Variants were constrained to the CpG sites of interest in each amplicon using the bcftools call function. A Perl script was used to generate a report of the percent methylation (of n = 1000 reads) at each CpG site, for each sample. 1000 sequence reads per sample were used to calculate the percentage of methylated CpG dinucleotides at each CpG site in each amplicon and plotted as histograms, with Lowess smoothing to generate trendlines. Deep sequence analysis of FRDA-DMR hypermethylation was displayed using a custom macro, which assigned rows for each of n = 300 sequencing reads and columns for each of the 11 CpG dinucleotides in amplicon 3 (which represented the FRDA-DMR), marking each coordinate black if methylated, and white if unmethylated. Individual reads (rows) were sorted with the most methylation stacked at the bottom. Epialleles were identified as follows: fully methylated with all 11 CpGs methylated, unmethylated with ≤ 2 CpGs methylated, and partially methylated with > 2 and < 11 CpGs methylated. Unmethylated epialleles were permitted to have up to 2 methylated CpGs because of the occasional and sporadic methylation seen in non-FRDA epialleles. Epiallele frequencies were calculated as percentages of the n = 300 individually sequenced DNA strands representing the FRDA-DMR.

Quality control and validation of bisulfite deep sequencing assay for FRDA-DMR methylation

Bisulfite conversion using either the Active Motif bisulfite conversion kit or the Zymo EZ DNA methylation kit in a side-by-side comparison of the same DNA samples resulted in $< 1.5\%$ standard deviation (SD), irrespective of high or relatively low methylation levels in the FRDA-DMR. Input genomic DNA of 0.1/0.5/1.0 µg for bisulfite conversion of the same DNA sample resulted in $< 1.5\%$ SD (we used 0.5 µg to generate our results). Reproducibility measurement, i.e. analysis of the same DNA sample in two separate sequencing assays, resulted in $< 3\%$ SD, using samples over a range of methylation levels. FRDA-DMR methylation analysis at depths of n = 100/300/1000 resulted in $< 2.5\%$ SD. With no statistical difference between n = 300 versus n = 1000 across multiple samples with widely varying methylation levels (Spearman correlation coefficient, ρ = 0.99), we used n = 300 to generate our FRDA-DMR data for enhanced multiplexing and efficient utilization of sequencing resources. The relatively homogenous population of FRDA patients with both expanded alleles containing > 500 repeats (n = 57 patients) were queried for any effect of age (at sample collection) or sex on FRDA-DMR methylation. Males and females showed similar levels of methylation (91% vs. 89%, not significant), and age showed minimal, if any, correlation (R^2 = 0.16).

Statistical analyses

Statistical tests were performed using GraphPad, Prism v8. Chi square and Fisher's exact tests were employed to compare proportions, as appropriate. Mann–Whitney was used to compare medians of non-normally distributed datasets where appropriate. Linear correlation was evaluated with Pearson correlation coefficient where appropriate, and the Benjamini and Hochberg method of false discovery rate was used to account for multiple comparisons. Second order polynomial (quadratic) correlations were performed, showing 95% confidence intervals. LASSO regression was performed to evaluate the relative roles of expanded GAA alleles (GAA1, GAA2, GAA1 + GAA2) and methylation in various regions (M1-M5, as depicted in Fig. 3C), using *FXN* transcript level as the response variable, because of the large number of covariates that may themselves be correlated. LASSO regression was done using R version 3.5.0 (R core team, 2018) and glmnet (version 2.0; ref. 52).

Supplementary Material

Supplementary Material is available at *HMG* online.

Conflict of Interest statement. None declared.

Funding

National Institutes of Health (R01 NS072418 to S.I.B., and R01 NS063856 to J.M.G.); Muscular Dystrophy Association (MDA604011 to S.I.B.); Friedreich's Ataxia Research Alliance (S.I.B.); the CureFA Foundation (S.I.B.)

References

1. Bidichandani, S.I. and Delatycki, M.B. (1993) [Updated 2017 Jun 1] Adam, M.P., Ardinger, H.H., Pagon, R.A., Wallace, S.E., Bean, L.J.H., Stephens, K. and Amemiya, A., Eds. Friedreich Ataxia. *GeneReviews*, Seattle (WA), https://www.ncbi.nlm.nih.gov/books/NBK1281/. (17 December 2020, date last accessed)
2. Campuzano, V., Montermini, L., Moltò, M.D., Pianese, L., Cossee, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A. *et al.* (1996) Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science*, **271**, 1423–1427.
3. Bidichandani, S.I., Ashizawa, T. and Patel, P.I. (1998) The GAA triplet-repeat expansion in Friedreich ataxia interferes with transcription and may be associated with an unusual DNA structure. *Am. J. Hum. Genet.*, **62**, 111–121.
4. Pianese, L., Turano, M., Lo Casale, M.S., De Biase, I., Giacchetti, M., Monticelli, A., Criscuolo, C., Filla, A. and Cocozza, S. (2004) Real time PCR quantification of frataxin mRNA in the peripheral blood leucocytes of Friedreich ataxia patients and carriers. *J. Neurol. Neurosurg. Psychiatry*, **75**, 1061–1063.

5. Chutake, Y.K., Lam, C., Costello, W.N., Anderson, M. and Bidichandani, S.I. (2014) Epigenetic promoter silencing in Friedreich ataxia is dependent on repeat length. *Ann. Neurol.*, **76**, 522–528.
6. Li, Y., Lu, Y., Polak, U., Lin, K., Shen, J., Farmer, J., Seyer, L., Bhalla, A.D., Rozwadowska, N., Lynch, D.R. *et al.* (2015) Expanded GAA repeats impede transcription elongation through the FXN gene and induce transcriptional silencing that is restricted to the FXN locus. *Hum. Mol. Genet.*, **24**, 6932–6943.
7. Campuzano, V., Montermini, L., Lutz, Y., Cova, L., Hindelang, C., Jiralerspong, S., Trottier, Y., Kish, S.J., Faucheux, B., Trouillas, P. *et al.* (1997) Frataxin is reduced in Friedreich ataxia patients and is associated with mitochondrial membranes. *Hum. Mol. Genet.*, **6**, 1771–1780.
8. Gervason, S., Larkem, D., Mansour, A.B., Botzanowski, T., Muller, C.S., Pecqueur, L., Le Pavec, G., Delaunay-Moisan, A., Brun, O., Agramunt, J. *et al.* (2019) Physiologically relevant reconstitution of iron-sulfur cluster biosynthesis uncovers persulfide-processing functions of ferredoxin-2 and frataxin. *Nat. Commun.*, **10**, 3566.
9. Fox, N.G., Das, D., Chakrabarti, M., Lindahl, P.A. and Barondeau, D.P. (2015) Frataxin accelerates [2Fe-2S] cluster formation on the human Fe-S assembly complex. *Biochemistry*, **54**, 3880–3889.
10. González-Cabo, P. and Palau, F. (2013) Mitochondrial pathophysiology in Friedreich's ataxia. *J. Neurochem.*, **126**(Suppl 1), 53–64.
11. Dürr, A., Cossee, M., Agid, Y., Campuzano, V., Mignard, C., Penet, C., Mandel, J.L., Brice, A. and Koenig, M. (1996) Clinical and genetic abnormalities in patients with Friedreich's ataxia. *N. Engl. J. Med.*, **335**, 1169–1175.
12. Filla, A., De Michele, G., Cavalcanti, F., Pianese, L., Monticelli, A., Campanella, G. and Cocozza, S. (1996) The relationship between trinucleotide (GAA) repeat length and clinical features in Friedreich ataxia. *Am. J. Hum. Genet.*, **59**, 554–560.
13. Montermini, L., Richter, A., Morgan, K., Justice, C.M., Julien, D., Castellotti, B., Mercier, J., Poirier, J., Capozzoli, F., Bouchard, J.P. *et al.* (1997) Phenotypic variability in Friedreich ataxia: role of the associated GAA triplet repeat expansion. *Ann. Neurol.*, **41**, 675–682.
14. Delatycki, M.B., Paris, D.B., Gardner, R.J., Nicholson, G.A., Nassif, N., Storey, E., MacMillan, J.C., Collins, V., Williamson, R. and Forrest, S.M. (1999) Clinical and genetic study of Friedreich ataxia in an Australian population. *Am. J. Med. Genet.*, **87**, 168–174.
15. Monrós, E., Moltó, M.D., Martínez, F., Canizares, J., Blanca, J., Vilchez, J.J., Prieto, F., de Frutos, R. and Palau, F. (1997) Phenotype correlation and intergenerational dynamics of the Friedreich ataxia GAA trinucleotide repeat. *Am. J. Hum. Genet.*, **61**, 101–110.
16. Regner, S.R., Wilcox, N.S., Friedman, L.S., Seyer, L.A., Schadt, K.A., Brigatti, K.W., Perlman, S., Delatycki, M., Wilmot, G.R., Gomez, C.M. *et al.* (2012) Friedreich ataxia clinical outcome measures: natural history evaluation in 410 participants. *J. Child Neurol.*, **27**, 1152–1158.
17. Metz, G., Coppard, N., Cooper, J.M., Delatycki, M.B., Durr, A., Di Prospero, N.A., Giunti, P., Lynch, D.R., Schulz, J.B., Rummey, C. *et al.* (2013) Rating disease progression of Friedreich's ataxia by the international cooperative ataxia rating scale: analysis of a 603-patient database. *Brain*, **136**, 259–268.
18. Tsou, A.Y., Paulsen, E.K., Lagedrost, S.J., Perlman, S.L., Mathews, K.D., Wilmot, G.R., Ravina, B., Koeppen, A.H. and Lynch, D.R. (2011) Mortality in Friedreich ataxia. *J. Neurol. Sci.*, **307**, 46–49.
19. Lecocq, C., Charles, P., Azulay, J.P., Meissner, W., Rai, M., N'Guyen, K., Pereon, Y., Fabre, N., Robin, E., Courtois, S. *et al.* (2016) Delayed-onset Friedreich's ataxia revisited. *Mov. Disord.*, **31**, 62–69.
20. Pousset, F., Legrand, L., Monin, M.L., Ewenczyk, C., Charles, P., Komajda, M., Brice, A., Pandolfo, M., Isnard, R., Tezenas du Montcel, S. *et al.* (2015) A 22-year follow-up study of long-term cardiac outcome and predictors of survival in Friedreich ataxia. *JAMA. Neurol*, **72**, 1334–1341.
21. Saveliev, A., Everett, C., Sharpe, T., Webster, Z. and Festenstein, R. (2003) DNA triplet repeats mediate heterochromatin-protein-1-sensitive variegated gene silencing. *Nature*, **422**, 909–913.
22. Herman, D., Jenssen, K., Burnett, R., Soragni, E., Perlman, S.L. and Gottesfeld, J.M. (2006) Histone deacetylase inhibitors reverse gene silencing in Friedreich's ataxia. *Nat. Chem. Biol.*, **2**, 551–558.
23. Greene, E., Mahishi, L., Entezam, A., Kumari, D. and Usdin, K. (2007) Repeat-induced epigenetic changes in intron 1 of the frataxin gene and its consequences in Friedreich ataxia. *Nucleic Acids Res.*, **35**, 3383–3390.
24. Al-Mahdawi, S., Pinto, R.M., Ismail, O., Varshney, D., Lymperi, S., Sandi, C., Trabzuni, D. and Pook, M. (2008) 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.
25. Rai, M., Soragni, E., Jenssen, K., Burnett, R., Herman, D., Coppola, G., Geschwind, D.H., Gottesfeld, J.M. and Pandolfo, M. (2008) HDAC inhibitors correct frataxin deficiency in a Friedreich ataxia mouse model. *PLoS One*, **3**, e1958.
26. Soragni, E., Herman, D., Dent, S.Y., Gottesfeld, J.M., Wells, R.D. and Napierala, M. (2008) Long intronic GAA*TTC repeats induce epigenetic changes and reporter gene silencing in a molecular model of Friedreich ataxia. *Nucleic Acids Res.*, **36**, 6056–6065.
27. Punga, T. and Buhler, M. (2010) Long intronic GAA repeats causing Friedreich ataxia impede transcription elongation. *EMBO Mol. Med.*, **2**, 120–129.
28. Kumari, D., Biacsi, R.E. and Usdin, K. (2011) Repeat expansion affects both transcription initiation and elongation in friedreich ataxia cells. *J. Biol. Chem.*, **286**, 4209–4215.
29. Kim, E., Napierala, M. and Dent, S.Y. (2011) Hyperexpansion of GAA repeats affects post-initiation steps of FXN transcription in Friedreich's ataxia. *Nucleic Acids Res.*, **39**, 8366–8377.
30. Chutake, Y.K., Costello, W.N., Lam, C. and Bidichandani, S.I. (2014) Altered nucleosome positioning at the transcription start site and deficient transcriptional initiation in Friedreich ataxia. *J. Biol. Chem.*, **289**, 15194–15202.
31. Chan, P.K., Torres, R., Yandim, C., Law, P.P., Khadayate, S., Mauri, M., Grosan, C., Chapman-Rothe, N., Giunti, P., Pook, M. *et al.* (2013) 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.
32. Li, Y., Polak, U., Bhalla, A.D., Rozwadowska, N., Butler, J.S., Lynch, D.R., Dent, S.Y.R. and Napierala, M. (2015) Excision of expanded GAA repeats alleviates the molecular phenotype of Friedreich's ataxia. *Mol. Ther.*, **23**, 1055–1065.
33. Soragni, E., Miao, W., Iudicello, M., Jacoby, D., De Mercanti, S., Clerico, M., Longo, F., Piga, A., Ku, S., Campau, E. *et al.* (2014) Epigenetic therapy for Friedreich ataxia. *Ann. Neurol.*, **76**, 489–508.

34. Silva, A.M., Brown, J.M., Buckle, V.J., Wade-Martins, R. and Lufino, M.M. (2015) Expanded GAA repeats impair FXN gene expression and reposition the FXN locus to the nuclear lamina in single cells. *Hum. Mol. Genet.*, **24**, 3457–3471.
35. Erwin, G.S., Grieshop, M.P., Ali, A., Qi, J., Lawlor, M., Kumar, D., Ahmad, I., McNally, A., Teider, N., Worringer, K. *et al.* (2017) Synthetic transcription elongation factors license transcription across repressive chromatin. *Science*, **358**, 1617–1622.
36. Castaldo, I., Pinelli, M., Monticelli, A., Acquaviva, F., Giacchetti, M., Filla, A., Sacchetti, S., Keller, S., Avvedimento, V.E., Chiariotti, L. *et al.* (2008) DNA methylation in intron 1 of the frataxin gene is related to GAA repeat length and age of onset in Friedreich ataxia patients. *J. Med. Genet.*, **45**, 808–812.
37. Evans-Galea, M.V., Carrodus, N., Rowley, S.M., Corben, L.A., Tai, G., Saffery, R., Galati, J.C., Wong, N.C., Craig, J.M., Lynch, D.R. *et al.* (2012) FXN methylation predicts expression and clinical outcome in Friedreich ataxia. *Ann. Neurol.*, **71**, 487–497.
38. Groh, M., Lufino, M.M., Wade-Martins, R. and Gromak, N. (2014) R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. *PLoS Genet.*, **10**, e1004318.
39. Li, L., Matsui, M. and Corey, D.R. (2016) Activating frataxin expression by repeat-targeted nucleic acids. *Nat. Commun.*, **7**, 10606.
40. Lai, J.I., Nachun, D., Petrosyan, L., Throesch, B., Campau, E., Gao, F., Baldwin, K.K., Coppola, G., Gottesfeld, J.M. and Soragni, E. (2019) Transcriptional profiling of isogenic Friedreich ataxia neurons and effect of an HDAC inhibitor on disease signatures. *J. Biol. Chem.*, **294**, 1846–1859.
41. Doi, A., Park, I.H., Wen, B., Murakami, P., Aryee, M.J., Irizarry, R., Herb, B., Ladd-Acosta, C., Rho, J., Loewer, S. *et al.* (2009) Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat. Genet.*, **41**, 1350–1353.
42. Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M. *et al.* (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.*, **41**, 178–186.
43. Pollard, S.M., Stricker, S.H. and Beck, S. (2009) Preview. A shore sign of reprogramming. *Cell Stem Cell*, **5**, 571–572.
44. Li, J., Li, Y., Wang, J., Gonzalez, T.J., Asokan, A., Napierala, J.S. and Napierala, M. (2020) Defining transcription regulatory elements in the human Frataxin gene: implications for gene therapy. *Hum. Gene Ther.*, **31**, 839–851.
45. Chutake, Y.K., Lam, C.C., Costello, W.N., Anderson, M.P. and Bidichandani, S.I. (2016) Reversal of epigenetic promoter silencing in Friedreich ataxia by a class I histone deacetylase inhibitor. *Nucleic Acids Res.*, **44**, 5095–5104.
46. Sharma, R., Bhatti, S., Gomez, M., Clark, R.M., Murray, C., Ashizawa, T. and Bidichandani, S.I. (2002) The GAA triplet-repeat sequence in Friedreich ataxia shows a high level of somatic instability in vivo, with a significant predilection for large contractions. *Hum. Mol. Genet.*, **11**, 2175–2187.
47. Timms, R.T., Tchasovnikarova, I.A. and Lehner, P.J. (2016) Position-effect variegation revisited: HUSHing up heterochromatin in human cells. *BioEssays*, **38**, 333–343.
48. Libri, V., Yandim, C., Athanasopoulos, S., Loyse, N., Natisvili, T., Law, P.P., Chan, P.K., Mohammad, T., Mauri, M., Tam, K.T. *et al.* (2014) Epigenetic and neurological effects and safety of high-dose nicotinamide in patients with Friedreich's ataxia: an exploratory, open-label, dose-escalation study. *Lancet*, **384**, 504–513.
49. Perdomini, M., Belbellaa, B., Monassier, L., Reutenauer, L., Messaddeq, N., Cartier, N., Crystal, R.G., Aubourg, P. and Puccio, H. (2014) Prevention and reversal of severe mitochondrial cardiomyopathy by gene therapy in a mouse model of Friedreich's ataxia. *Nat. Med.*, **20**, 542–547.
50. Piguet, F., de Montigny, C., Vaucamps, N., Reutenauer, L., Eisenmann, A. and Puccio, H. (2018) Rapid and complete reversal of sensory ataxia by gene therapy in a novel model of Friedreich ataxia. *Mol. Ther.*, **26**, 1940–1952.
51. de Jonge, H.J., Fehrmann, R.S., de Bont, E.S., Hofstra, R.M., Gerbens, F., Kamps, W.A., de Vries, E.G., van der Zee, A.G., te Meerman, G.J. and ter Elst, A. (2007) Evidence based selection of housekeeping genes. *PLoS One*, **2**, e898.
52. Friedman, J., Hastie, T. and Tibshirani, R. (2010) Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.*, **33**, 1–22.