Sizing Pathogenic *FXN* GAA Repeat Expansions with Electronic Genome Mapping Using the Nabsys OhmX™ Platform and RepX™ Analysis Pipeline

Executive Summary

- Electronic genome mapping (EGM) and the RepX™ Analysis for Repeat Expansion Disorders (RepX Analysis) pipeline were evaluated for the detection and sizing of *FXN* GAA repeat expansions.
- Four control, three carrier, and three affected reference samples were analyzed and compared to gel-based sizing as an orthogonal method.
- RepX-derived repeat counts were concordant with gel-derived size ranges, enabling clear classification of pathogenic and non-pathogenic alleles.
- Quantitative sizing of GAA expansion repeats was achieved using native, double-stranded high molecular weight (HMW) DNA.
- *FXN* repeat expansion detection and sizing using EGM is part of a broader multi-locus workflow in which more than 16 repeat expansion loci can be assessed simultaneously in a single run.

Introduction

More than 50 genetic disorders are caused by repeat expansion mutations, in which short nucleotide motifs expand beyond normal ranges and disrupt gene function. These disorders include Friedreich's ataxia (FA), Fragile X syndrome, and myotonic dystrophy, among others.

As in some other genetic repeat expansion disorders, *FXN* repeat length correlates broadly with disease severity and age of onset, making accurate sizing essential for both diagnostic confirmation and research applications.[1] Accurate detection and sizing of expanded repeats remains technically challenging, particularly for large expansions spanning hundreds to thousands of repeat units.

Long-range PCR and repeat-primed PCR are the most common assays for detecting and sizing *FXN* repeat expansions.[2] These methods are locus-specific and effective for many short tandem repeat expansions; however, they are limited to targeted analysis of a single gene at a time and PCR conditions and primers must be carefully tailored for each expansion. Very large alleles may amplify inefficiently or fail to amplify entirely, leading to underestimation or misclassification.[3]

Short-read sequencing cannot reliably span long repetitive tracts and therefore cannot directly measure large intronic repeat expansions.[4] Long-read sequencing approaches can detect repeat expansions but typically require targeted enrichment and may be affected by amplification bias or reduced accuracy in long, low-complexity regions.

PCR remains the most common confirmatory test for *FXN* expansions. Yet, different PCR methods may produce discordant size estimates, reinforcing the need for orthogonal approaches for validation.

EGM performed on the Nabsys OhmX™ Platform provides an amplification-free method for assessing genomic interval lengths using native double-stranded HMW DNA.[5–7] When coupled with the Nabsys RepX Analysis pipeline, interval differences corresponding to repeat expansion can be measured quantitatively, generating repeat counts, confidence intervals, and probabilities of exceeding pathogenic thresholds. A detailed description of the computational framework was previously provided in the *FMR1* repeat expansion technical note.[8] The RepX Analysis pipeline is part of a broader multi-locus workflow capable of simultaneously characterizing more than 16 repeat expansion loci from a single EGM run; this technical note focuses on its application to the *FXN* locus.

This technical note evaluates the performance of EGM and the RepX Analysis pipeline for detecting and sizing *FXN* GAA repeat expansions across control, carrier, and affected reference samples. Results are compared to gel-based measurements of the same DNA sample to assess concordance and suitability for orthogonal validation. Comparisons to literature values for the samples can be erroneous because the expansion repeats are unstable and change length over time.

Friedreich's Ataxia

FA is an autosomal recessive neurodegenerative disorder caused by the expansion of a GAA trinucleotide repeat within intron 1 of the *FXN* gene on chr9. Normal alleles typically contain 5–33 repeat units, while pathogenic alleles contain 66 or more repeats and often expand to several hundred or over one thousand units.[4,9] An intermediate range of approximately 34–65 repeats has been reported, though it is rare and may or may not be associated with clinical manifestations.

The *FXN* gene encodes frataxin, a nuclear-encoded mitochondrial protein involved in iron-sulfur cluster biogenesis and cellular iron homeostasis.[4,10] Reduced frataxin expression leads to mitochondrial dysfunction and progressive neurodegeneration. Clinical manifestations typically begin between ages 10 and 15 and may include progressive ataxia, cardiomyopathy, diabetes or impaired glucose tolerance, and musculoskeletal abnormalities.[11] In 2023, omaveloxolone became the first FDA-approved therapy for FA.[12]

The estimated prevalence of FA is approximately 0.5 per 100,000 individuals, with carrier frequency estimated between 1 in 60 and 1 in 100 in certain populations.[4, 13]

Results

FA Sample Evaluation and Orthogonal Confirmation by PCR Gel Electrophoresis

Three carrier and three affected cell lines (cell IDs GM15847, GM16219, GM16237, and GM16216, GM16243, GM16798, respectively) were obtained from the NIGMS Human Genetic Cell Repository at the Coriell Institute for Medical Research (**Table 1**). As reported by previous studies, repeat counts from Coriell do not always align with repeat size ranges determined by PCR/gel electrophoresis.[14] For this evaluation, gel-based sizing of the same DNA sample was used as the primary orthogonal comparator.

The PCR primers and protocol are described in the Appendix. Gel electrophoresis of PCR products was used to confirm the presence and approximate sizing range of expanded alleles across the cohort. The gel image also provides context for samples exhibiting broader sizing ranges (consistent with mosaic allele size variability), which can be challenging for single-value reporting by some methods (**Figure 1**).

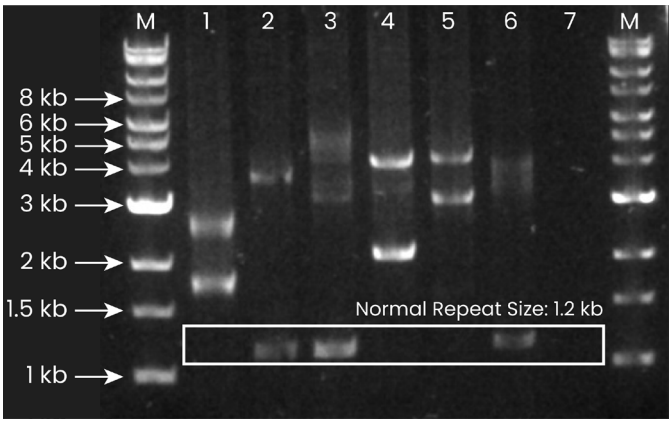

Figure 1. PCR gel confirming FXN GAA repeat expansion sizes. PCR amplification products spanning the *FXN* GAA repeat region were resolved by gel electrophoresis. Lane order (left to right): M: Marker, 1: GM16216, 2: GM16219, 3: GM16237, 4: GM16798, 5: GM16243, 6: GM15847, 7: Negative Control, M: Marker. The New England Biolabs 1 kb Extend DNA Ladder was used to estimate allele size. Band migration patterns correspond to normal, carrier, and expanded alleles, where the normal alleles are 1,192 bp with the primers used (white box). Broader band distributions in some samples reflect variability in repeat size.

Cell ID	Genotype	Number of Repeats (Coriell)	Number of Repeats (Gel)
GM16216	Affected	200 and 500	109-275 and 275-609
GM16219	Carrier	< 66 and 570	0-109 and 609-942
GM16237	Carrier	< 66 and 700	0-109, 609-942, and 942-1609
GM16798	Affected	750 and 1000	275-609 and 942-1275
GM16243	Affected	670 and 1170	609-942 and 942-1275
GM15847	Carrier	< 66 and 760	0-109, 609-942, and 942-1275

Table 1. Coriell cell lines evaluated for *FXN* GAA repeat expansion. Coriell cell lines analyzed in this study, including reported repeat counts from Coriell and repeat size ranges determined by PCR gel electrophoresis. Gel-derived ranges were used as the primary orthogonal comparator for assessing concordance with EGM and RepX pipeline results.

Gel-based sizes were estimated by subtracting the normal allele size generated by the PCR primers used (1,192 bp) from the marker sizes that the band fell between, and then dividing by three to account for the trinucleotide repeats.

EGM Data Generation with the OhmX Platform and RepX Analysis Pipeline

HMW DNA was prepared from cultured cell lines using the New England Biolabs HMW DNA Extraction Kit for Cells and Blood and the standard OhmX genome preparation workflow. Briefly, frozen pellets (1×10^6 cells) were processed to isolate ultra-long HMW DNA, nicked with Nt.BspQI and Nb.BssSI, labeled, coated with RecA, and analyzed on a 256-channel OhmX Detector to generate single-molecule interval maps spanning the whole genome including the *FXN* repeat region. Two injections were analyzed per sample as replicates, but only a single injection was necessary to evaluate the number of repeats present.

The RepX Analysis pipeline models candidate repeat interval lengths at the *FXN* locus and assigns repeat-spanning molecules to the best-supported allele hypotheses. For each allele, the RepX Analysis pipeline reports the observed repeat count, repeat-spanning coverage, probability of exceeding the pathogenic threshold (≥ 66 repeats), and a 95% confidence interval, along with additional metrics (data not shown). A more detailed description of the computational modeling framework is provided in the *FMR1* technical note.[8]

Control samples (cell IDs GM12891, GM12892, HG003, and HG004) without pathogenic expansion were analyzed alongside the FA cohort and classified as non-pathogenic, with observed repeat counts falling within normal ranges and with low probabilities of exceeding the pathogenic threshold. Carrier samples demonstrated the separation of one normal allele and one expanded allele. Expanded alleles showed a high probability of pathogenic expansion, while normal alleles remained below the threshold. Observed repeat counts were concordant with gel-based ranges across replicate injections. Affected samples carrying two expanded alleles were classified as pathogenic for both alleles. The RepX Analysis pipeline reported probabilities of full mutation near 1.0 for each modeled allele. Confidence intervals were reproducible across injections and essentially aligned with gel-defined repeat ranges.

A summary of repeat counts, confidence intervals, spanning coverage, and pathogenic probabilities across all evaluated samples (control, carrier, and affected) is provided below (**Table 2**).

Repeat-spanning coverage influenced confidence interval width but did not alter pathogenic classification when more than 20 spanning molecules were available for both alleles. Across all samples, the RepX Analysis pipeline outputs were essentially concordant with gel-derived sizing.

Disease Status	Sample	Repeat Annotation (by gel)	Run #	Genome Coverage (X)	Allele A				Allele B			
					Observed Repeat Count	Repeat coverage	P(Full Mutation)	95% C.I.	Observed Repeat Count	Repeat coverage	P(Full Mutation)	95% C.I.
Control Samples (No expansion)	HG003	< 66	1	98	22	25	0.04	[0,63]	22	24	0.05	[0,63]
			2	84	30	19	0.11	[0,76]	30	20	0.11	[0,75]
	GM12891	< 66	1	102	7	17	0.03	[0,60]	7	17	0.03	[0,60]
			2	98	13	23	0.03	[0,57]	13	23	0.03	[0,57]
	GM12892	< 66	1	53	1	22	0.01	[0,49]	1	22	0.01	[0,49]
			2	66	18	19	0.05	[0,65]	18	20	0.05	[0,64]
	HG004	< 66	1	127	4	33	0.00	[0,42]	4	33	0.00	[0,42]
			2	81	1	16	0.03	[0,58]	1	17	0.02	[0,56]
Carriers (1 allele > 66 repeats)	GM15847	< 66, 609-942, & 942-1275	1	70	963	12	1.00	[893,1077]	25	19	0.08	[0,72]
			2	120	808	37	1.00	[759,858]	11	23	0.02	[0,55]
	GM16219	< 66 & 609-942	1	82	916	20	1.00	[848,987]	0	26	0.00	[0,43]
			2	91	872	33	1.00	[819,926]	0	27	0.00	[0,43]
	GM16237	< 66, 609-942, & 942-1609	1	98	1278	27	1.00	[1214,1345]	32	24	0.09	[0,72]
			2	100	1174	32	1.00	[1116,1234]	42	25	0.17	[3,84]
Affected (2 alleles > 66 repeats)	GM16216	109-275 & 275-609	1	95	498	20	1.00	[438,560]	242	26	1.00	[195,292]
			2	63	619	15	1.00	[547,695]	286	17	1.00	[226,348]
	GM16243	609-942 & 942-1275	1	87	1001	14	1.00	[917,1088]	529	13	1.00	[454,608]
			2	93	1122	25	1.00	[1058,1189]	661	28	1.00	[607,716]
	GM16798	275-609 & 942-1275	1	57	1041	14	1.00	[956,1129]	329	21	1.00	[274,386]
			2	79	1011	22	1.00	[944,1080]	307	25	1.00	[257,359]

Table 2. RepX-derived repeat counts, confidence intervals, and pathogenic probabilities for evaluated samples. The RepX Analysis pipeline output summary for control, carrier, and affected samples. For each run, reported values include the observed repeat count (maximum *a posteriori* estimate), repeat-spanning molecule coverage, the probability that the repeat count exceeds the pathogenic threshold (≥ 66 repeats), and the 95% confidence interval.

Repeat-Length Variability and Mosaicism

Some samples demonstrated broad gel-based sizing ranges consistent with variability in repeat length (**Figure 1**). The RepX Analysis pipeline's waterfall plots provide a molecule-level view of per-allele repeat size distributions at specific loci. For the representative example shown in **Figure 2**, data from three injections were combined to increase repeat-spanning coverage (combined genome coverage of 226x), enabling clearer visualization of repeat-length heterogeneity. In GM16237, the waterfall plot at the *FXN* locus shows two distinct allele populations: Allele A molecules (green) form a steep gradient with a staircase-like distribution above the pathogenic threshold, while Allele B molecules (blue) cluster near or below the normal range. The distribution shape of Allele A repeat lengths is consistent with somatic instability or mosaicism at this locus, while the distribution shape of Allele B repeat lengths is consistent with somatic stability.

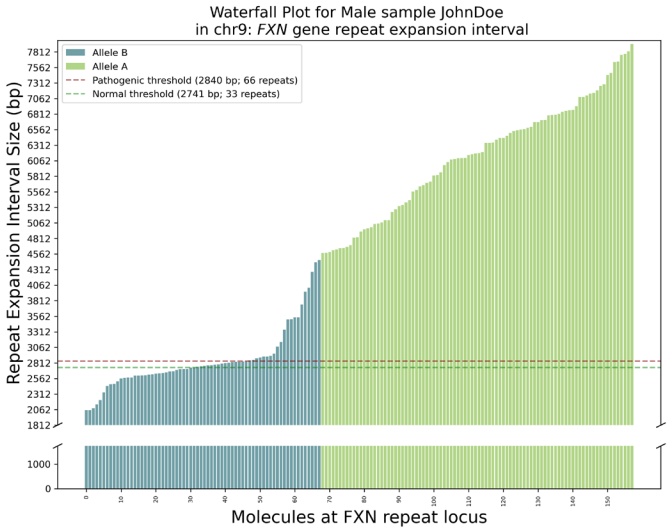

Figure 2. RepX waterfall plot resolves two distinct *FXN* allele populations in carrier sample GM16237, revealing repeat-length heterogeneity in the expanded allele. Data from three combined injections (226x total genome coverage) were analyzed using the RepX Analysis pipeline. Each bar represents a single molecule spanning the FXN repeat locus, ranked by repeat count. Allele A (green) molecules span a wide range of repeat lengths well above the pathogenic threshold (dashed line), consistent with somatic instability or mosaicism of the expanded allele. Allele B (blue) molecules cluster around a central value, consistent with somatic stability. In this example, we see both expanded and normal molecules.

Unlike PCR and gel electrophoresis, which allow visualization of only the most prominent alleles, the RepX Analysis pipeline repeat size distribution plots allow visualization of all alleles present in a sample. Furthermore, EGM is not subject to the same size biases as PCR in that long alleles are detected as effectively as short alleles due to the absence of amplification in the protocol. While formal mosaicism detection is still under development in current computational workflows, molecule-level waterfall plots and allele modeling outputs can provide evidence of repeat heterogeneity and support qualitative assessment in research settings.

Summary

EGM on the OhmX Platform, combined with the RepX Analysis pipeline, accurately detected and sized *FXN* GAA repeat expansions across control, carrier, and affected reference samples. Observed repeat counts were concordant with PCR gel-derived size ranges and enabled clear discrimination of pathogenic and non-pathogenic alleles.

This study supports the use of EGM as an orthogonal method for repeat expansion confirmation and demonstrates the applicability of the RepX Analysis pipeline for quantitative analysis of intronic GAA repeat expansions using native double-stranded HMW DNA.

References

1. Dürr A, Cossee M, Agid Y, et al. Clinical and genetic abnormalities in patients with Friedreich's ataxia. *N Engl J Med*. 1996;335(16):1169-1175. doi:10.1056/NEJM199610173351601
2. Lucain M, Duffourd Y, Malbos M, Vitobello A, Thauvin-Robinet C, Thomas Q. Past, present, and future of genomic technologies in cerebellar ataxias. *J Neurol*. 2025;272(12):753. doi:10.1007/s00415-025-13478-2
3. Park J, Dufke C, Fleszar Z, et al. Long-read sequencing identifies mosaic sequence variations in Friedreich's ataxia-GAA repeats. *Int J Mol Sci*. 2025;26(11):4969. doi:10.3390/ijms26114969
4. Bidichandani SI, Delatycki MB, Napierala M, Duquette A. Friedreich ataxia. In: Adam MP, Bick S, Mirzaa GM, Pagon RA, Wallace SE, Amemiya A, eds. *GeneReviews®*. University of Washington, Seattle; 1993. Accessed February 13, 2026. http://www.ncbi.nlm.nih.gov/books/NBK1281/
5. Kaiser MD, Davis JR, Grinberg BS, et al. Automated structural variant verification in human genomes using single-molecule electronic DNA mapping. *bioRxiv*. Preprint posted online May 22, 2017:140699. doi:10.1101/140699
6. Oliver JS, Catalano A, Davis JR, et al. High-definition electronic genome maps from single molecule data. *bioRxiv*. Preprint posted online May 18, 2017:139840. doi:10.1101/139840
7. Thompson JF, Oliver JS. Mapping and sequencing DNA using nanopores and nanodetectors. *Electrophoresis*. 2012;33(23):3429-3436. doi:10.1002/elps.201200136
8. Nabsys 2.0, LLC. Pathogenic *FMR1* repeat expansions confirmed with EGM data analyzed using the Nabsys RepX pipeline. Published online October 8, 2025. https://www.nabsys.com/resources/pathogenic-*fmr1*-repeat-expansions-confirmed-with-egm-data-analyzed-using-the-nabsys-repx-pipeline
9. Campuzano V, Montermini L, Moltò MD, et al. Friedreich's ataxia: Autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science*. 1996;271(5254):1423-1427. doi:10.1126/science.271.5254.1423
10. Pastore A, Puccio H. Frataxin: A protein in search for a function. *J Neurochem*. 2013;126 Suppl 1:43-52. doi:10.1111/jnc.12220
11. Delatycki MB, Bidichandani SI. Friedreich ataxia-pathogenesis and implications for therapies. *Neurobiol Dis*. 2019;132:104606. doi:10.1016/j.nbd.2019.104606
12. Lynch DR, Hauser L, McCormick A, et al. Randomized, double-blind, placebo-controlled study of interferon-γ 1b in Friedreich ataxia. *Ann Clin Transl Neurol.* 2019;6(3):546-553. doi:10.1002/acn3.731
13. Buesch K, Zhang R. A systematic review of disease prevalence, health-related quality of life, and economic outcomes associated with Friedreich's ataxia. *Curr Med Res Opin.* 2022;38(10):1739-1749. doi:10.1080/03007995.2022.2112870
14. Kim E. *Definition of the landscape of chromatin structure at the frataxin gene in Friedreich's ataxia.* The University of Texas Graduate School of Biomedical Sciences at Houston; 2011. https://digitalcommons.library.tmc.edu/utgsbs_dissertations/206
15. Lamont PJ, Davis MB, Wood NW. Identification and sizing of the GAA trinucleotide repeat expansion of Friedreich's ataxia in 56 patients. Clinical and genetic correlates. *Brain J Neurol.* 1997;120 (Pt 4):673-680. doi:10.1093/brain/120.4.673
16. Long A, Napierala JS, Polak U, et al. Somatic instability of the expanded GAA repeats in Friedreich's ataxia. *PloS One.* 2017;12(12):e0189990. doi:10.1371/journal.pone.0189990

Appendix

PCR Protocol

The following PCR primers were used[15]

Primer 1: CAA TCC AGG ACA GTC AGG GCT TT

Primer 2: GGG ATT GGT TGC CAG TGC TTA AAA GTT AG

To set up the PCR reaction mixture, TaKaRa LA Taq DNA Polymerase with GC Buffer I was used, along with 200 ng of DNA, according to the manufacturer's instructions.

PCR was performed using the following protocol:[16]

1. Initial Denaturing: 94°C 3 min
2. 20 cycles:
   a. Denaturing: 94°C 20 sec
   b. Annealing: 64°C 30 sec
   c. Extension: 68°C 5 min
3. 9 cycles, where the anneal/extend step increases by 15 sec each cycle:
   a. Denaturing: 94°C 20 sec
   b. Anneal/Extend: 68°C 5 min
4. Final Extension: 68°C 7 min

Find additional OhmX Platform resources at nabsys.com/applications/repx.