Solving amplification of high-GC expanded repeats to enable rare disease detection using long-read sequencing.

Manuel Rozas, Jennifer Hsieh, Javier Gimpel, Javier Cáceres, Catalina Valdivia, Tamara Akentjew, Isidora Arriagada, Isa Marx.
Kura Biotech Spa, Puerto Varas, Región de Los Lagos, Chile.

KURA® biotech

www.kurabiotech.com

INTRODUCTION

Long-read sequencing (LRS) is the gold standard for resolving complex genomic regions, yet clinical applications with limited sample input often require targeted amplification to generate sufficient material and recover full genomic structures. This is particularly critical for repeat expansion disease, such as Amyotrophic Lateral Sclerosis(ALS), Friedreich's Ataxia (FrA), Fragile X Syndrome (FrX) and Huntington´s Disease (HD).

OBJECTIVE

To develop a high-fidelity amplification method, using DeCodi-Fi™ products, to robustly enrich long, GC-rich, and repetitive genomic regions while preserving full sequence integrity for precision long-read sequencing.

METHODS

Sample selection: Genomic DNA from unaffected controls and affected individuals (sourced from the Coriell Institute) was used to evaluate the amplification of repeat expansions within the *C9orf72* (ALS), *FXN* (FrA), *FMR1* (FrX), and *IT15* (HD) genes. Human *BRCA1* (8,4 kb, 23,6 kb and 36,4 kb) and *SMN1* targets (28.2 kb) were amplified to evaluate long-range PCR performance across polymerases.

PCR protocol: Platinum SuperFi II (Thermo Fisher Scientific) and Q5 Hot Start High-Fidelity DNA Polymerase (New England Biolabs) were evaluated according to their respective manufacturer-recommended buffers, enhancers, and cycling protocols.
DeCodi-Fi™ and its specialized formulations were optimized for specific targets: DeCodi-Fi™ High-Fidelity PCR kit, DeCodi-Fi™ Long&Complex, and DeCodi-Fi™ High-GC Repeat Enhancer.

Cycling protocols were optimized for specific primer pairs and targets using stepdown PCR, ramp rate control, and wobble temperatures, and applied uniformly across all polymerases.

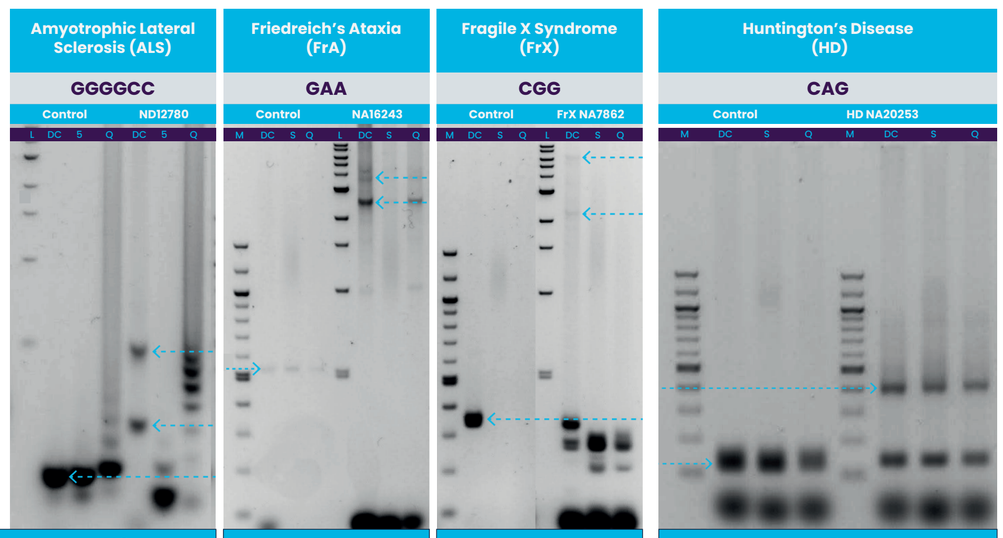
RESULTS

Amplification method of Disease-Associated repeat targets for long-read sequencing

Selecting the optimal Polymerase for challenging repeats

We evaluated several high-fidelity polymerases to identify the most effective system for four major repeat expansions: ALS, FrA, FrX, and HD. We focused on the enzymes' ability to overcome the biochemical barriers, such as extreme GC-content and secondary structures, that typically cause standard PCR failure.

A)



B)

	ALS		FrA		FrX		HD	
Template DNA (Expected size)	**Control** (143 bp)	**ND12780** (178/466 bp)	**Control** (520 bp)	**NA16243** (2510/4010 bp)	**Control** (350 bp)	**NA7862** (2339bp/ >3239bp)	**Control** (105 bp)	**NA20253** (119/353 bp)
DeCodi-Fi™ products	✓	✓	✓	✓ (unspecific)	✓	✓ (non reproducible)	✓	✓
SuperFi II	✓	✗	✓	✗	✗	✗	✓	✓
Q5	✗ (larger)	✗	✓	✓ (weak board)	✗	✗	✓	✓

Figure 1. A) Amplification of disease-associated repeats (ALS, FrA, FrX, HD) using DeCodi-Fi™ products(DC), SuperFi II (S), and Q5 (Q). PCR products from **unaffected control** and **affected individuals** were analyzed by agarose gel electrophoresis. All enzymes were tested under equivalent conditions, with GC buffers or enhancers added as needed. For targets composed exclusively of long GC-rich repeats, such as ALS (GGGGCC) and FrX (CGG), the DeCodi-Fi™ High-GC Repeats Enhancer kit was required. For long expansions >500 bp with moderate GC content, such as FrA (GAA), the DeCodi-Fi™ Long&Complex formulation was used. Finally, for targets with lower GC content and shorter lengths (under 500 bp), such as HD (CAG), the standard DeCodi-Fi™ PCR Kit provided optimal amplification. Light blue arrows indicate detected products. **M:** 100 bp ladder; **L:** 1 kb ladder.

B) **Summary of amplification outcomes shown in Figure 1.** ✓ indicates successful amplification; ✗ indicates failure. DeCodi-Fi™ amplified all templates, including the 6.5 kb mosaic allele in FrX, typically missed due to its presence in a small cell subpopulation.

Verifying sequence integrity via long-read sequencing

Using Huntington's Disease (HD) as a model, we compared how these amplified products perform on the Oxford Nanopore (ONT) platform. This step confirmed that the enrichment process preserved the full sequence structure and motif accuracy needed for precision analysis.

Table 1. Median motif counts and sequencing metrics for unaffected control and a Huntington's disease sample NA20253 for Allele 1(A1)/Allele 2(A2), comparing DeCodi-Fi™ (DC), SuperFi II (S), and Q5 (Q). The motif count refers to the median number of repeat motifs (CAG) detected per allele. On-target indicates the total number of sequencing reads aligned to the repeat region. Spanning reads represent the number of reads that fully span the repeat region for each allele.

Genome	Expected motif count (A1/A2)	Enzyme	Median motif count (A1/A2)	On-target	Spanning reads (A1/A2)
Unaffected	<36	DC	17/17	9958	2367/2067
		S	18/18	9948	2104/2254
		Q	18/18	8122	1856/1675
HD NA20253	22/100	DC	22/103	8930	2652/230
		S	22/106	8016	2526/380
		Q	22/104	5706	1766/279

DeCodi-Fi™ was identified as the optimal system by successfully resolving all expansions, achieving robust **amplification in high-GC regions (up to 79.26%) and recovering challenging targets such as the 6.5 kb *FMR1* ([CGG]>1000) allele.**

Long-read sequencing confirmed that DeCodi-Fi™ amplification **preserved repeat structure and accurate motif counts,** with a high number of spanning reads.

Optimization for complex clinical targets (FrX and FrA)

To ensure reliability across diverse clinical cases, we refined our protocols for Fragile X and Friedreich's Ataxia. By testing various patient genotypes with different expansion lengths, we optimized the method to handle the most challenging samples.

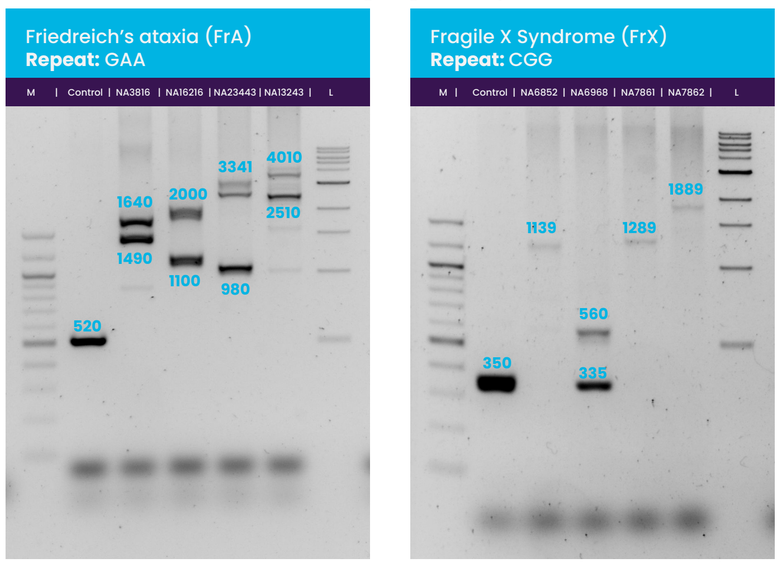


Figure 2. Amplification of disease-associated repeats (FrA and FrX) of unaffected control and different disease affected samples using DeCodi-Fi™ products. PCR products were analyzed by agarose gel electrophoresis. For FrA affected samples the DeCodi-Fi™ Long&Complex was required. For FrX affected samples the DeCodi-Fi™ with High-GC Repeats Enhancer kit was required. **M:** 100 bp ladder; **L:** 1 kb ladder.

Optimized protocols reliably **amplified difficult expansions across different patient samples.**

Amplification of large-scale human targets (*BRCA1* and *SMN1*)

Following the identification of the optimal system for repeat expansions, we evaluated the long-range high-fidelity PCR performance of three polymerases on large-scale human targets: Breast Cancer 1 (*BRCA1*) and Survival Motor Neuron 1 (*SMN1*).

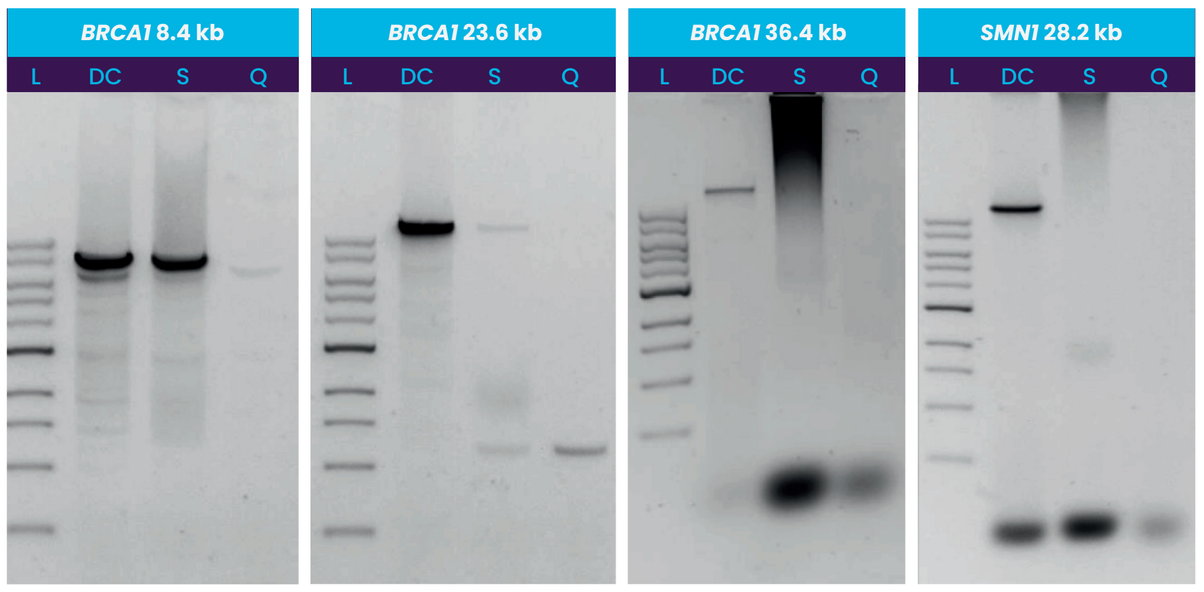


Figure 3. Long-range amplification of clinically relevant targets: *BRCA1* and *SMN1*. Amplification of *BRCA1* (8.4 kb, 23.6 kb, 36.4 kb) and *SMN1* (28.2 kb) using DeCodi-Fi™ products **(DC)**, SuperFi II **(S)**, and Q5 **(Q)**. **L:** 1 kb ladder. For most targets, the DeCodi-Fi™ High-Fidelity PCR kit with High-Fidelity Buffer was utilized and successfully amplified all targets Note that the 36.4 kb *BRCA1* fragment required the DeCodi-Fi™ High-Fidelity PCR kit supplemented with an additional enhancer.

DeCodi-Fi™ successfully **amplified human DNA fragments up to 36.4 kb** ensuring locus continuity for large-scale structural variant analysis.

CONCLUSIONS

We have developed a high-fidelity amplification method using the DeCodi-Fi™ portfolio that robustly enriches long, GC-rich, and repetitive regions. The method successfully resolved all four disease-associated expansions, including the 6.5 kb (79.26% GC) Fragile X, while extending amplification capabilities to 36.4 kb (compared to 8.4 kb with alternative enzymes). By ensuring higher coverage and increased spanning reads, this approach facilitates the characterization of clinically relevant, challenging regions. These results demonstrate that DeCodi-Fi™ preserves sequence integrity across complex loci, providing a reliable tool for precision long-read sequencing applications.

Learn more about DeCodi-Fi™

Download the poster here