Investigating the hypermethylation and triplet-repeat expansion of Friedreich's Ataxia using PCR-free targeting

Both hypermethylation and repeat expansion are found in intron 1 of the frataxin locus in Friedreich's Ataxia, and are associated with low levels of gene transcription and with disease symptoms

Contact: **publications@nanoporetech.com** More information at: **www.nanoporetech.com** and **publications.nanoporetech.com**

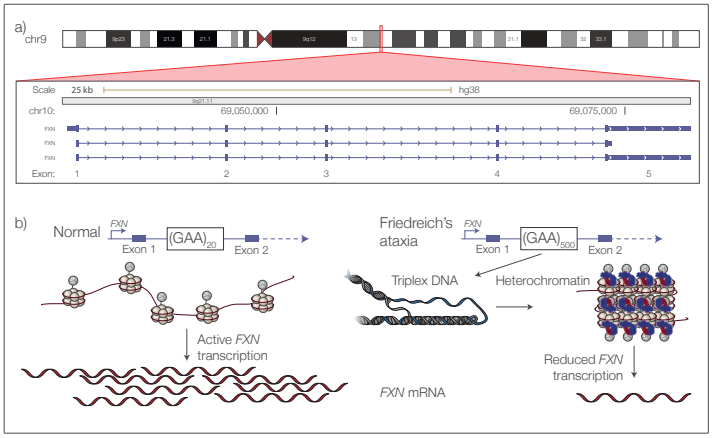

Fig. 1 The *FXN* locus a) location on chromosome 9 b) possible disease mechanisms

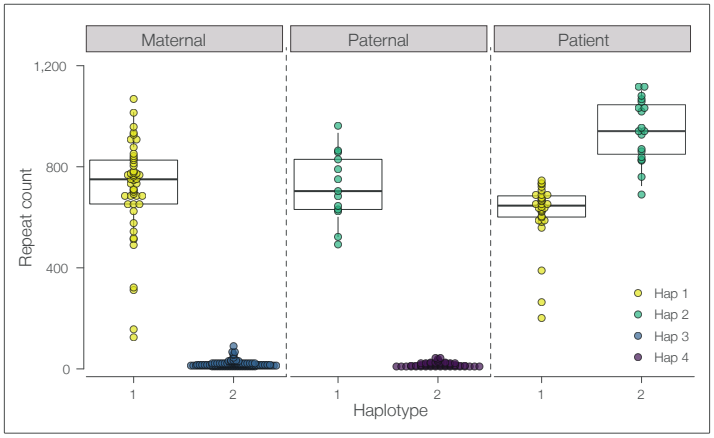

Fig. 2 GAA repeat length in carrier parents and their affected child

Friedreich's Ataxia is a common, autosomal recessive genetic disease

Friedreich's Ataxia is a recessive, progressive neurodegenerative movement disorder affecting around 1 in 50,000 people. Sufferers of the disease produce too little of the protein frataxin, coded by the *FXN* gene on chromosome 9 (Fig. 1a). Frataxin is involved in the assembly of iron-sulphur clusters in mitochondria. The intron between exons 1 and 2 of the gene contains a GAA triplet, which is typically below 20 copies in length, but in carriers this can expand to over a thousand repeats. The disease mechanism is not fully understood, but is thought to involve reduced levels of *FXN* transcription, possibly arising from unusual DNA conformations caused by the repeat or by hypermethylation and heterochromatin formation near to the repeat (Fig. 1b).

Measuring the length of the unstable GAA repeat expansion in base space

The presence of a long, low-complexity repeat and the presence of hypermethylation mean that a PCR-based targeting approach would not be suitable for investigating the genetics of the disease. We therefore enriched the *FXN* locus from carrier parents and their affected child using Cas9. We aligned the resulting reads to the human reference genome hg38 using *minimap2* and haplotyped with *MarginPhase*. We counted repeat length in base-space (Fig. 2). Sequencing results agree well with the values derived from Southern blotting. A high level of somatic instability has been reported in expanded *FXN* alleles, accounting for the difference in repeat size between the father and child.

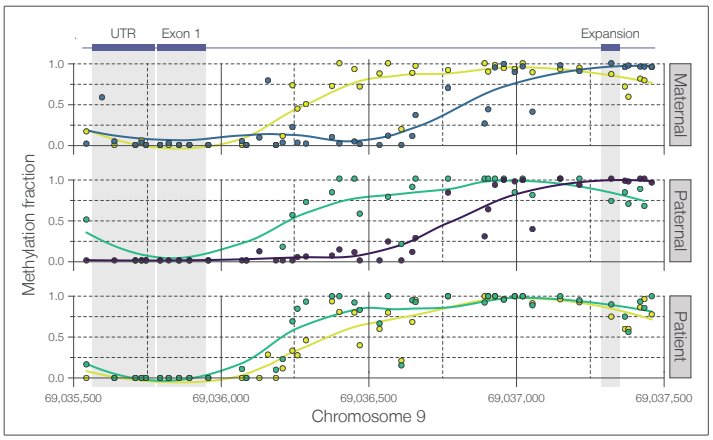

Fig. 3 Methylation of intron 1 at the *FXN* locus in carrier parents and their affected child

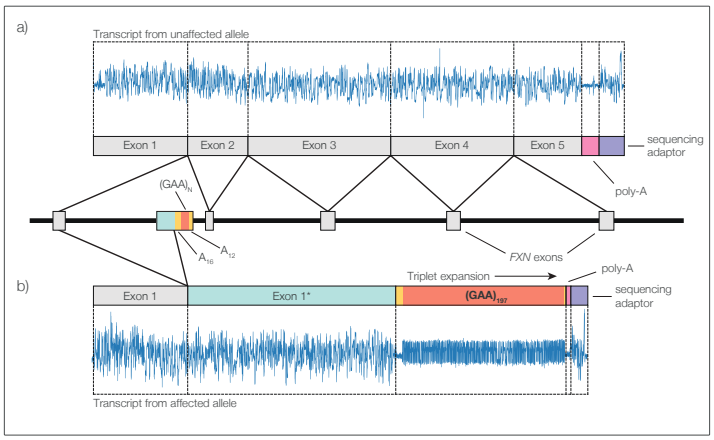

Fig. 4 Transcription at the *FXN* locus from a) unaffected and b) affected alleles

Detecting hypermethylation in intron 1, upstream of the triplet repeat

PCR-free targeting means that epigenetic modifications are retained in the DNA strands, and these can be detected in the raw nanopore sequencing signal. We used the tool *Nanopolish* to call CpG methylation across the *FXN* locus in the carrier and patient samples. In both the maternal and paternal carrier samples, reads from the repeat-expanded haplotype were hyper-methylated in the ~2 kb region between exon 1 and the repeat expansion, whereas in the unaffected haplotypes no hypermethylation was detected in the region (Fig. 3 top and middle panels). In contrast, in the patient sample, both haplotypes were expanded, and both were hypermethylated. The haplotype with the larger expansion showed a slightly higher level of methylation (Fig. 3 bottom panel).

Direct cDNA sequencing of *FXN* transcripts shows irregular splicing in a carrier parent

To investigate the effect of repeat expansion and hypermethylation on *FXN* transcription we prepared whole-transcriptome direct cDNA libraries from a carrier parent sample. As expected, the read count of *FXN* transcripts was lower than for a non-carrier. Interestingly, two populations of transcript were detected. The more abundant population consisted of regularly-spliced transcripts (Fig. 4a) corresponding to the unaffected allele, but the minor population consisted of irregularly spliced RNAs (Fig. 4b), as has been reported in other triplet-repeat expansion diseases. The triplet is present in the transcript from the affected allele as part of an alternative exon. It can also be seen that the length of the triplet is lower than expected (Fig. 4b), implying that the low complexity repeat may present difficulties for RNA polymerase II.