SCA4: Correlation of age of onset and *ZFHX3* repeat length

JOEL WALLENIUS MSc[1], LUDMILA KAPLUN PhD[2], GREICE KRAUTZ-PETERSON PhD[2], ALEXANDER KAPLUN PhD[2], ANDREAS PUSCHMANN MD PhD[1]

[1]Dept. for Neurology, Lund University, Skåne University Hospital, Sweden [2]Variantyx Inc. Framingham MA, USA

Background

We and others have recently reported that exonic GCC-trinucleotide repeat expansions in *ZFHX3* are the cause of spinocerebellar ataxia 4 (SCA4, **Table**).

Our aim with the present study was to determine the repeat lengths in five Swedish SCA4 families and to correlate them with age at onset, and to compare with recently published repeat lengths from other SCA4 families.

First authors	Publication date	Families	Long-read sequencing method; DNA source	Reference
Wallenius J, Kafantari E, et al.	29 Nov 2023 (Epub)	• 5 families from southern Swedish Skåne Region	Pacific Biosciences (PacBio); blood	Am J Hum Genet. 2024;111:82-95.
Chen Z, Gustavsson EK, et al.	10 Jan 2024 (Epub)	• large kindred of Swedish ancestry in Utah* • patient from another family with Swedish ancestry, from Iowa	Oxford Nanopore Technology (ONT); lymphoblastoid cell lines	Mov Disord. 2024;39:486-497.
Figueroa KP, Gross C, et al.	29 April 2024 *(medRxiv preprint 28 Oct 2023)*	• large kindred of Swedish ancestry in Utah* • large Northern German kindred • patients from 6 additional German kindreds	PacBio or ONT; blood	Nat Genet. 2024 Apr 29. doi:10.1038/s41588-024-01719-5.
Paucar M, Nilsson D, et al.	*3 Oct 2023 (preprint)*	• 3 Swedish families, individual age at onset not reported	ONT; cell culture (1 patient)	*medRxiv (not peer reviewed)* *https://doi.org/10.1101/2023.10.03.23296230*

Table: Published families with SCA4 and *ZFHX3* repeat expansions. Information as provided in the publications. *Two groups simultaneously analyzed the original Utah SCA4 kindred (Flanigan 1996).

Methods

For this study, we performed Oxford Nanopore Technology (ONT) long-read sequencing using DNA from blood samples of 11 SCA4 patients from 5 southern Swedish families. We had previously performed Pacific Biosciences (PacBio) long-read sequencing on two SCA4 patients. We retrieved short-read, long-read and age-at-onset data as provided in the previous reports (**Table**) and our own data. We correlated age at onset with repeat lengths and examined the measurements of repeat lengths using the different sequencing technologies.

Results

Long-read single-molecule sequencing generated reads spanning the entire length of non-expanded and expanded alleles. The exact length of the expanded alleles differed slightly between the reads from different DNA molecules in the blood samples, likely indicating somatic mosaicism in blood cells and marked instability of the repeats (**Figure 1**). Non-expanded repeats were uniform in length and showed interruptions. Interruptions were lost in expanded alleles. We have previously hypothesized that loss of the interruptions may represent the initial founder event, and that loss of all interruptions increases the likelihood for expansion of the repeat (**Figure 2**).

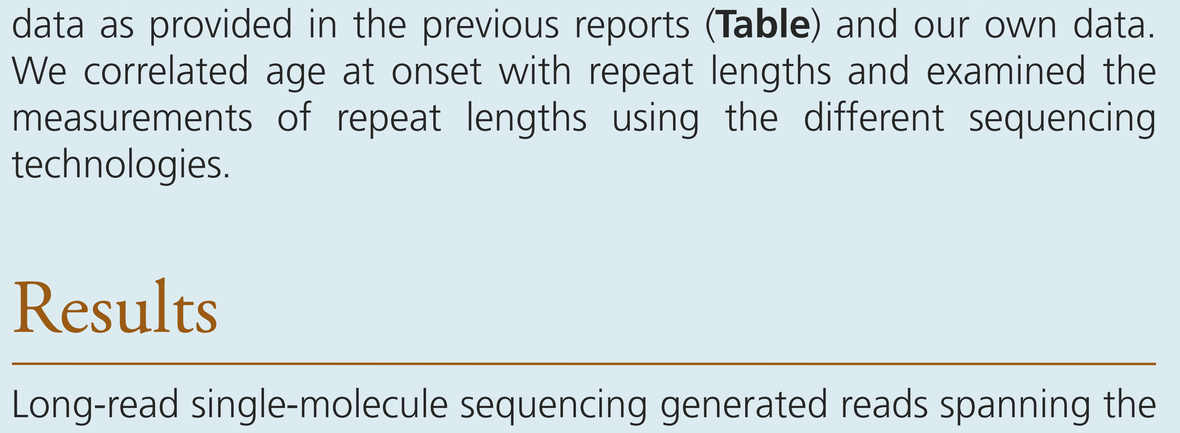

Figure 1: Representative examples of Oxford Nanopore single-molecule long-read sequencing results from two SCA4 patients. Expanded alleles are shown on the top and non-expanded alleles on the bottom of each image.

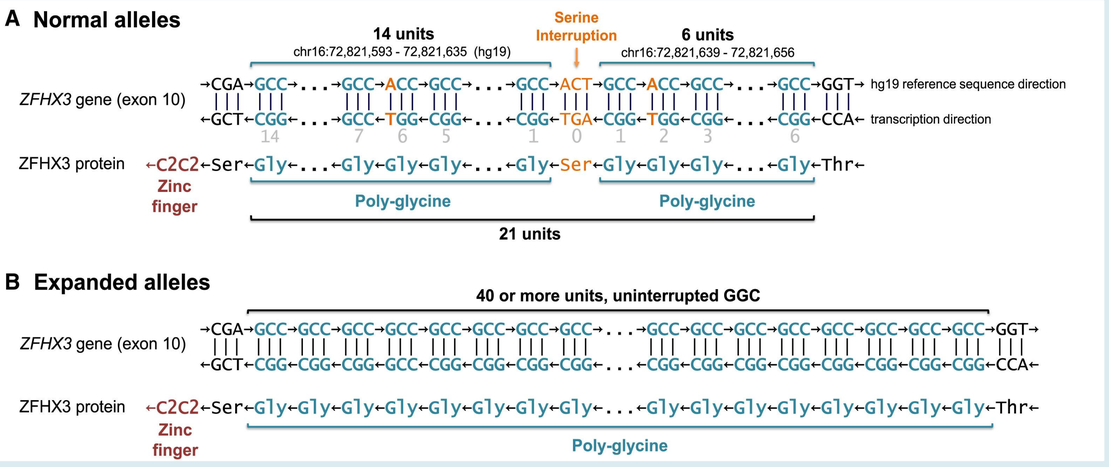

Figure 2: Schematic representation of non-expanded *ZFHX3* GCC repeats, with interruptions, and expanded repeats in SCA4 patients where interruptions were lost.
From: Wallenius J et al. Am J Hum Genet. 2024 Jan 4;111(1):82-95.

Lengths of the expanded repeats correlated inversely with age of onset of SCA4 symptoms; correlation was not perfect (r=-0.86; **Figure 3**). There was the problem that expansions vary from molecule to molecule even in blood samples, and presumably much more between blood and different brain areas or peripheral nerves, which hampers exact correlation.

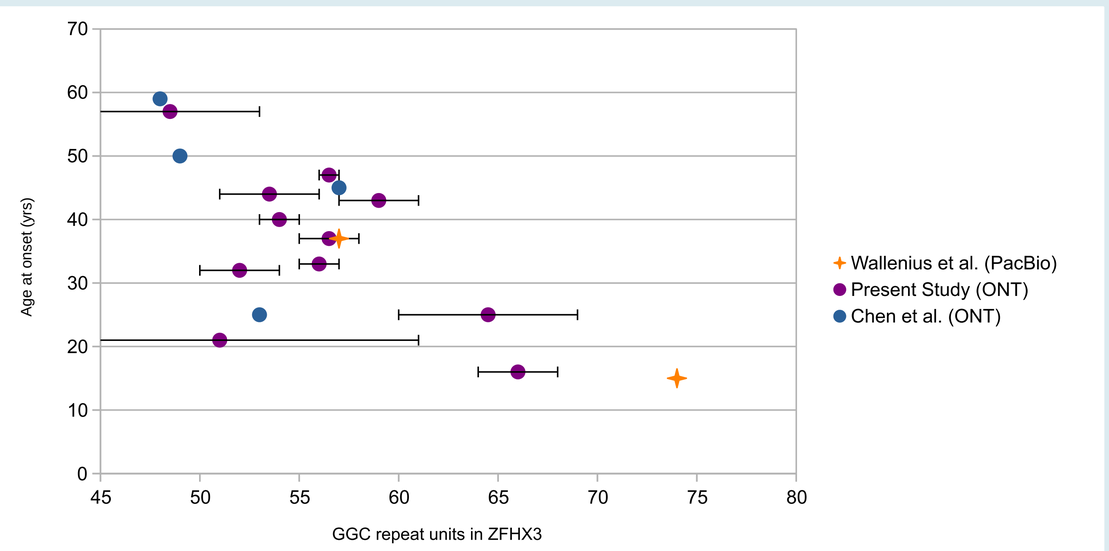

Figure 3: Correlation of *ZFHX3* repeat length and age of onset in SCA4 patients. Recently, Figueroa et al. published a similar inverse correlation in their SCA4 patient series (r = –0.797, see Table for reference). For the data from the present study, we provide the range of repeat expansion length as measured by ONT long-read sequencing, probably caused by somatic mosaicism between blood cells.

Conclusions

- Short-read sequencing (Illumina) can identify *ZFHX3* expansions, but long-read sequencing is required to determine the repeat length.
- Longer GCC repeats in *ZFHX3* were correlated with earlier disease onset.
- Long-read sequencing confirmed our previous finding that expanded *ZFHX3* alleles have lost interruptions that commonly occur in non-expanded alleles.
- Single-molecule sequencing revealed small differences in repeat lengths in blood samples, likely indicating instability of the repeat lengths and somatic mosaicism between blood cells.
- This implies that repeat lengths in brain or peripheral nerves might differ from those observed in blood samples.
- Somatic instability with mosaicism and loss of interruptions are known mechanisms from other neurological repeat disorders. The instability of the expanded repeats offers possibilities to develop treatments for SCA4, as is being pursued for other repeat expansion disorders.

Contact

- Andreas Puschmann, Professor of Neurology, Skåne University Hospital, Lund, Sweden and Department of Clinical Sciences, Lund University, Sweden: **andreas.puschmann@med.lu.se**
- Alexander Kaplun PhD, Vice President Product, Variantyx Inc., Framingham MA, USA: **alex.kaplun@variantyx.com**