Rapid diagnosis of SCA36 in a three-generation family using short-read whole genome sequencing data

Haloom Rafehi 1,2,3, David J. Szmulewicz 4,5, Kate Pope 6, Mathew Wallis 7,8, John Christodoulou 9,10, Susan M White 10,11,12, Martin B Delatycki 6,10,11, Paul J Lockhart 6,10, Melanie Bahlo* 1,2

1. Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia
2. Department of Medical Biology, University of Melbourne, Melbourne, Australia
3. Epilepsy Research Centre, University of Melbourne, Austin Health, Melbourne, Australia
4. Cerebellar Ataxia Clinic, Neuroscience Department, Alfred Health, Melbourne, VIC 3004, Australia;
5. Balance Disorders and Ataxia Service, Royal Victorian Eye & Ear Hospital, East Melbourne, VIC 3002, Australia
6. Bruce Lefroy Centre for Genetic Health Research, Murdoch Children's Research Institute, 50 Flemington Road Parkville, Victoria, 3052, Australia
7. Tasmanian Clinical Genetics Service, Tasmanian Health Service, Tasmania, Australia
8. School of Medicine and Menzies Institute for Medical Research, University of Tasmania, Tasmania, Australia
9. Brain and Mitochondrial Research Group, Murdoch Children's Research Institute, 50 Flemington Road Parkville, Victoria, 3052, Australia
10. Department of Pediatrics, The University of Melbourne, 30 Royal Parade Parkville, Victoria, 3052, Australia
11. Victorian Clinical Genetics Services, 50 Flemington Road Parkville, Victoria, 3052, Australia
12. Murdoch Children's Research Institute, 50 Flemington Road Parkville, Victoria, 3052, Australia

*Corresponding author: Melanie Bahlo

Keywords: ataxia, repeat expansions, short tandem repeats, diagnosis, exSTRa

Financial disclosure/conflicts of interest: Nothing to report.

Funding sources for study:
The funding bodies had no role in the design, analysis, interpretation or reporting of results.

Abstract

Background: Spinocerebellar ataxias (SCA) are often caused by expansions of short tandem repeats (STRs). Recent methodological advances have made repeat expansion (RE) detection with whole genome sequencing (WGS) feasible.
Objectives: To determine the genetic basis of ataxia in a large, multigenerational Australian pedigree, with autosomal dominant inheritance and possible anticipation.
Methods and Results: WGS was performed on three affected relatives. The sequence data was screened for known pathogenic REs using three repeat expansion detection tools: exSTRa, ExpansionHunter and GangSTR. This screen provided a clear and rapid diagnosis (<5 days from receiving the sequencing data) of SCA36, a very rare form of ataxia which is caused by an intronic hexanucleotide GGCCTG RE in *NOP56*.
Conclusions: We have demonstrated that diagnosis of rare ataxias caused by REs is highly feasible and cost effective with WGS. We propose a change in current clinical practice, such that WGS be implemented as the frontline, cost effective methodology for molecular testing of individuals with a clinical diagnosis of ataxia.

Introduction

Spinocerebellar ataxias (SCA) comprise a group of rare, progressive neurological disorders that can be caused by deleterious point mutations or pathogenic expansions of short tandem repeats (STRs). Ataxias caused by repeat expansions (REs) of STRs are difficult to diagnose, with low throughput methods such as locus-specific sizing or repeat primed PCR assays being most commonly performed [1]. Clinical testing is routinely available for SCAs caused by more common REs such as Friedreich's ataxia, SCA1, 2, 3, 6 and 7, but not rarer forms such as CANVAS, SCA 8, 10, 12, 17, 31, 36 and 37. Hence molecular diagnosis of ataxias due to REs, particularly less common REs such as those underlying SCA36 and SCA37, can be hard to obtain. In addition, current RE tests are expensive and can be time consuming, sometimes requiring international shipping of biospecimens or DNA. This has multiple negative outcomes, including uncertainty and a potential diagnostic odyssey for patients, while clinicians and health providers are unable to offer accurate prognosis, counselling and best clinical care.

Methodological advances in bioinformatics in the last two years have made the simultaneous testing for all known REs through whole genome sequencing (WGS) feasible [2] . Additionally, multiple novel REs have been discovered in neurological disorders using WGS in the last two years alone, including in epilepsy [3] and ataxia [4-6]. Indeed, there is a growing recognition of the importance of REs and potential contribution to unsolved neurogenetic disease [7].

In a recent study, we utilised WGS to discover a novel AAGGG RE in the gene encoding replication factor 1 (*RFC1*) that causes cerebellar ataxia with neuropathy and bilateral vestibular areflexia syndrome (CANVAS) [4]. We thus provided a genetic CANVAS diagnosis to patients from 18 families in our Australian CANVAS cohort. Additionally, we also used the same WGS data to re-diagnose patients who tested negative for the *RFC1* RE with other forms of ataxia, including one individual with SCA3, also caused by a RE.

In this study, we present a case study in which a three generation Australian pedigree with autosomal dominant ataxia was rapidly diagnosed with SCA36 using WGS-based bioinformatic approaches, and subsequently validated by two independent international service providers. SCA36 is a very rare ataxia with slow progression, and is characterized by adult-onset gait and appendicular ataxia, oculomotor abnormalities, cerebellar dysarthria and hyperreflexia. A definite clinical diagnosis of SCA36 is difficult due to overlap with other ataxias, and therefore targeted testing of this locus is not routinely considered. This issue is compounded by the fact that in Australia and many other countries there is no accredited diagnostic testing available for SCA36. A small number of international service providers do offer single locus SCA36 testing, but the test is expensive, turnaround times can be long, and cannot provide the exact size of the RE. This study highlights that accredited diagnostic WGS, which is available in many countries, can provide an alternative single test to rapidly diagnose SCA36 and other unsolved ataxias, effectively addressing a previously unmet clinical need.

Results

A large Australian family of European ancestry with seventeen individuals affected by ataxia over three generations was recruited to this study (Figure 1). Segregation is consistent with autosomal dominant inheritance, with apparently complete penetrance. At least 18 members of this family in 4 generations present with ataxia. Onset of symptoms is in mid adulthood and is slowly progressive. Lifespan is not significantly reduced with a number of affected individuals living into their 80s. A detailed neurological examination was undertaken on the three subjects and found that all displayed an ataxic gait, characterized by a broadened base of support, lateral veering, reduced walking speed, irregular foot placement, an increased stance phase and increased foot rotation angles [8, 9]. Oculomotor abnormalities were consistent with cerebellar impairment and included saccadic horizontal and vertical visual pursuits [10] and vestibulo-ocular reflex suppression [11]; and dysmetric saccades to target [12]. Cerebellar dysarthria with consequent reduction in speech intelligibility was also present [13, 14]. Appendicular ataxia was seen in all four limbs and manifested by dysmetria [15, 16] and intention tremor [15, 17]. Global upper and lower limb hyperreflexia was also present on examination [18].

WGS was performed for three affected individuals from the family (Figure 1) and the data was analysed for a pre-defined list of 12 candidate repeat expansions known to cause ataxia (SCA1,2,3,6,7,8,10,12,17,36, Friedreich's ataxia and CANVAS) using exSTRa [19] and ExpansionHunter [20]. This analysis identified the SCA36 RE in all three individual's WGS data, with none of the 167 controls exhibiting a RE. Based on evidence from the repeat read distributions plotted by exSTRa the RE was present in a heterozygous state in all affected individuals sequenced. SCA36 is caused by an expansion of the intronic hexanucleotide GGCCTG RE in *NOP56* (Figure 2A, OMIM #614153). Expansion Hunter was used to estimate the size of the expanded allele determining a lower bound of ~800 repeat units (Figure 2B). This is a lower bound due to the limitations of sizing of very large expansions with short-read sequencing, with a known bias towards estimating fewer repeat motifs. This is a significant distinction as the pathogenic threshold for SCA36 has previously been described as greater than 650 repeats [21]. Individuals with normal repeat size display a motif range of 1-12 repeat units. More precise sizing of SCA36 RE requires a Southern blot, however no such Southern blot is currently available commercially to our knowledge due to the rarity of this RE. Visualisation of the locus in IGV confirms a large RE present in all three individuals (Figure 2C).

Subsequent to identification of SCA36 in the family, we performed confirmatory linkage analysis using genotypes extracted from the WGS data with LINKDATAGEN [22]. While underpowered, this analysis identified 15 regions of identity by descent with LOD scores between 1 and 1.5 (data not shown). One of these regions (chr20:257608-7368304, LOD=1.5024, 20p13) overlapped SCA36, and also encompasses the genes for SCA23 (*PDYN*) and SCA35 (*TGM6*). A second linkage region on chromosome 4 intersects with the gene that causes SCA41. No potentially pathogenic sequence variants or RE were identified in these genes in the three affected individuals.

Examination of clinical features for this family confirmed features consistent with SCA36. To confirm the diagnosis, molecular diagnostic SCA36 testing was ordered. As diagnostic testing is not currently available in Australia, diagnosis was sought through two independent international service providers (Fulgent and MNG Laboratories). Both returned a positive result, identifying one allele as greater than 70 repeats. However, only alleles >650 are considered full-penetrance and diagnostic of SCA36, while alleles between 15 and 650 repeats are considered variants of uncertain significance. Neither provider offers Southern blot or any other test to accurately size the RE.

Discussion

SCA36 is a rare condition that shares considerable clinical overlap with other ataxias, meaning molecular diagnostic testing is difficult to access. Here, we report for the first time that WGS can provide rapid and efficient diagnosis of SCA36. This is of significant clinical importance, given the difficulty and lack of availability of diagnostic testing for SCA36 worldwide, with affected individuals remaining underdiagnosed for this genetic mutation. Importantly, other than Southern blot, which is rarely used, no other test can identify >600 repeats for the SCA36 locus. We demonstrate that WGS can offer a rapid and efficient method for ataxia caused by REs, such as SCA36.

SCA36 is a rare monogenic form of ataxia that includes motor neuron involvement and is caused by an intronic GGCCTG (AGGCCC) hexanucleotide repeat expansion in *NOP56* (OMIM #614153). It was first identified in patients of Japanese ancestry, using linkage mapping in three unrelated families [21]. Nine unrelated affected individuals were identified in a large Japanese ataxia cohort (n=251). These individuals shared a common haplotype, suggesting a single founder event for SCA36 in the Japanese population. Individuals with SCA36-mediated ataxia have since been identified in China, the US, Italy and Spain, with the latter country demonstrating a relatively high incidence of SCA36, owing to a founder effect [23]. It appears likely that there are multiple founder events for SCA36 RE as has been observed for other repeat expansions, such as Huntington disease [24] and SCA3 [25].

SCA36 is exceptionally rare - in the initial Japanese study which described SCA36, the pathogenic allele was detected in 3.6% of patients in an unsolved ataxia cohort of 251 patients [21]. In a recent US study of 577 ataxia patients, SCA36 was detected in only 0.7% of unsolved ataxia patients [26]. Therefore, routine screening of ataxia patients for SCA36 using RP-PCR or by Southern blot is not considered practical or cost effective. We propose that changing clinical practice to utilise WGS as a stand-alone frontline test is a potentially cost-effective and efficient diagnostic pathway for individuals with a clinical diagnosis of ataxia. This method allows for the screening of all known RE simultaneously while also allowing examination of point mutations in known ataxia genes. Furthermore, the WGS data can subsequently be used for novel gene discovery in unsolved cases.

The rapidity and ease of genetic diagnosis for this family with WGS-based RE detection analysis, after decades of genetic testing, will extrapolate to other rare REs as well. In the ataxias more

broadly, numerous diagnostic procedures such as MRI, nerve conduction studies and cognitive testing are performed, in part to inform diagnostic genetic testing. However our experience, and globally, is that despite these extensive investigations, a diagnosis is only achieved in at best ~30% of cases [27-29]. The diagnostic journey is long (up to 20 years) and often unsuccessful. We strongly advocate the implementation of WGS for molecular diagnostic testing of individuals with a clinical diagnosis of ataxia. To facilitate these analyses we maintain up to date lists of known RE for all currently used bioinformatic RE detection tools. These can be obtained from https://github.com/bahlolab/Bio-STR-exSTRa.

Methods

Institutional Ethics Committee approval was provided by the Royal Children's Hospital (Melbourne, HREC 28907), and written informed consent was obtained from all participants prior to study. Clinical details were collected from clinical assessments and review of medical records. Genomic DNA was isolated from blood using standard techniques and WGS was performed fee for service by the Australian Genome Research Facility. Libraries were prepared using a PCR free kit (TruSeq) and sequenced using a Novaseq (Illumina). Alignment and haplotype calling were performed based on the GATK best practice pipeline. All samples were aligned to the hg19 reference genome using BWA-mem. Duplicate marking, local realignment, and recalibration were performed with GATK. Samples were screened for a short-list of 12 STRs known to cause ataxia: SCA1,2,3,6,7,8,10,12,17,36, Friedreich's ataxia and CANVAS. RE detection tools exSTRa and ExpansionHunter were used, using default parameters, with additional parameters for ExpansionHunter: inclusion of off-target reads and read-depth of 30 and min-anchor-mapq of 20. Results were compared to 167 control WGS samples from the GTEx project.
https://raw.githubusercontent.com/bahlolab/exSTRa/master/inst/extdata/repeat_expansion_disorders.txt

Acknowledgements

We acknowledge contributions from Michael Hildebrand, Natasha Brown and Samuel Berkovic. P.J.L is supported by the Vincent Chiodo Foundation. M.B. was supported by an NHMRC Senior Research Fellowship (GNT1102971). D.J.S is supported by an NHMRC Early Career Fellowship (1129595). Whole genome sequencing costs were funded by Austin Health and the Victorian Medical Research Acceleration Fund through the Austin Health Undiagnosed Diseases Program. This work was supported in part by the Victorian Government's Operational Infrastructure Support Program and the NHMRC Independent Research Institute Infrastructure Support Scheme (IRIISS).

Figures

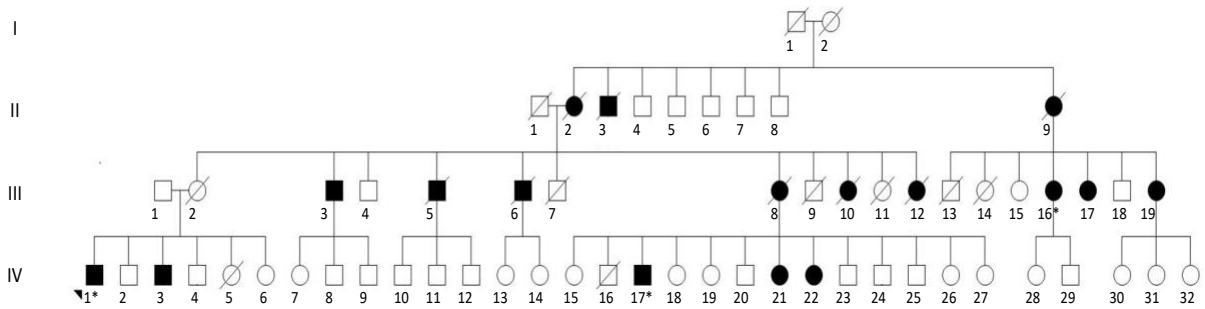

Figure 1. Multigenerational ataxia family pedigree. Individuals with confirmed ataxia have been shaded in black.

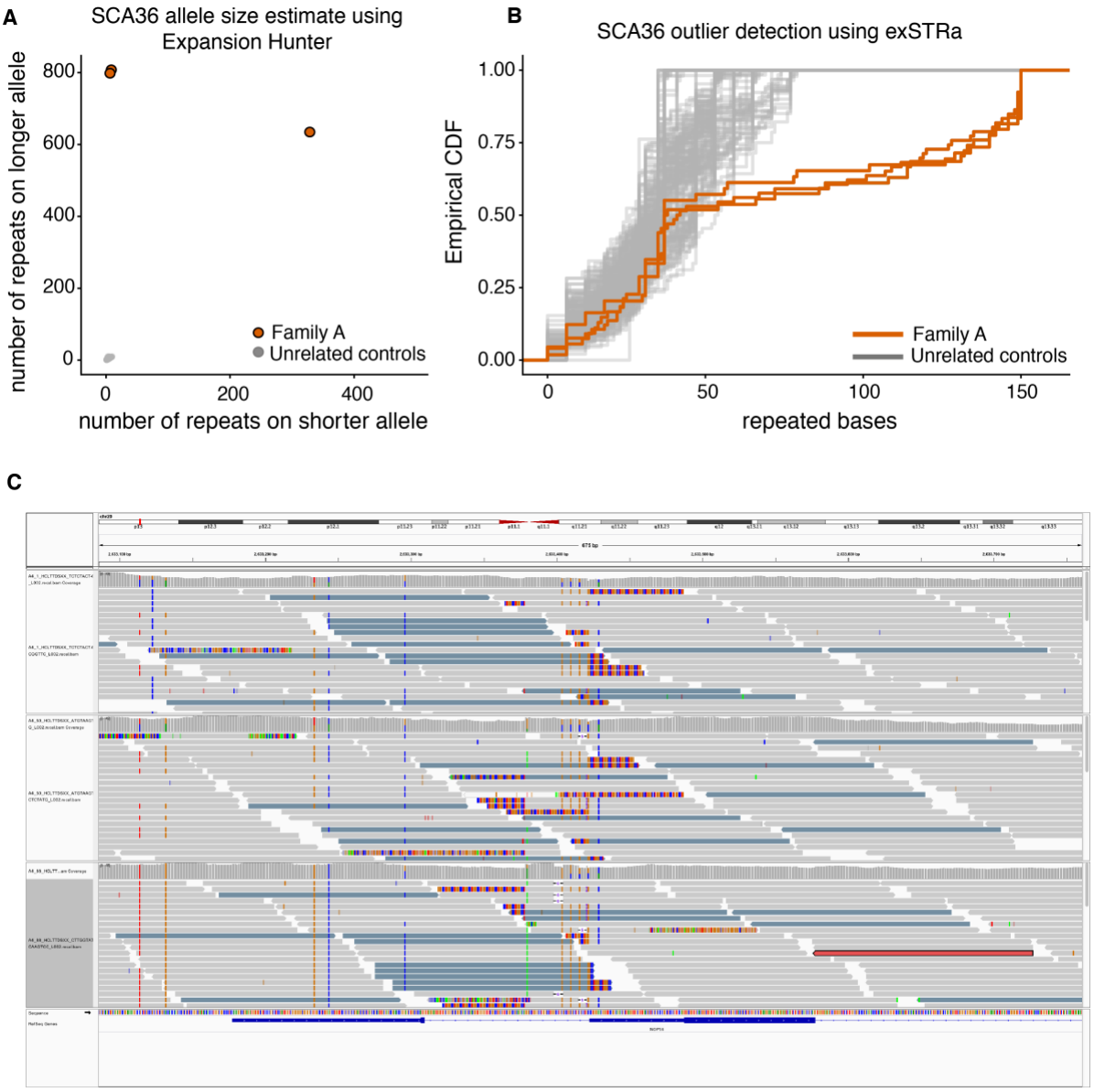

Figure 2. Identification of the SCA36 repeat expansion in a large family with ataxia.

Allele size estimate was performing using ExpansionHunter (A) while outlier testing was performed using exSTRa (B). Orange points/lines indicate the three family affected family members. The grey points/lines are the 167 controls samples. (C) Visualisation in IGV of the intronic hexanucleotide GGCCTG RE in the gene *NOP56* in the three patients. The blue-grey reads represent reads where the mate contains the expanded GGCCTG motif, which has mismapped to a GGCCTG-rich region of the genome (chr5:145,838,637-145,838,6).

References

1. de Silva, R.N., et al., *Diagnosis and management of progressive ataxia in adults.* Pract Neurol, 2019. **19**(3): p. 196-207.
2. Bahlo, M., et al., *Recent advances in the detection of repeat expansions with short-read next-generation sequencing.* F1000Res, 2018. **7**.
3. Ishiura, H., et al., *Expansions of intronic TTTCA and TTTTA repeats in benign adult familial myoclonic epilepsy.* Nat Genet, 2018. **50**(4): p. 581-590.
4. Rafehi, H., et al., *Bioinformatics-Based Identification of Expanded Repeats: A Non-reference Intronic Pentamer Expansion in RFC1 Causes CANVAS.* Am J Hum Genet, 2019. **105**(1): p. 151-165.
5. Cortese, A., et al., *Biallelic expansion of an intronic repeat in RFC1 is a common cause of late-onset ataxia.* Nat Genet, 2019. **51**(4): p. 649-658.
6. van Kuilenburg, A.B.P., et al., *Glutaminase Deficiency Caused by Short Tandem Repeat Expansion in GLS.* N Engl J Med, 2019. **380**(15): p. 1433-1441.
7. Lohmann, K. and N. Bruggemann, *Rediscovery of repeat expansions: Solving the unsolved cases.* Mov Disord, 2019. **34**(9): p. 1300.
8. Stolze, H., et al., *Typical features of cerebellar ataxic gait.* J Neurol Neurosurg Psychiatry, 2002. **73**(3): p. 310-2.
9. Buckley, E., C. Mazza, and A. McNeill, *A systematic review of the gait characteristics associated with Cerebellar Ataxia.* Gait Posture, 2018. **60**: p. 154-163.
10. Thier, P. and U.J. Ilg, *The neural basis of smooth-pursuit eye movements.* Curr Opin Neurobiol, 2005. **15**(6): p. 645-52.
11. Chambers, B.R. and M.A. Gresty, *The relationship between disordered pursuit and vestibulo-ocular reflex suppression.* J Neurol Neurosurg Psychiatry, 1983. **46**(1): p. 61-6.
12. Sharpe, J.A., A.W. Lo, and H.E. Rabinovitch, *Control of the saccadic and smooth pursuit systems after cerebral hemidecortication.* Brain, 1979. **102**(2): p. 387-403.
13. Ackermann, H., et al., *Speech deficits in ischaemic cerebellar lesions.* J Neurol, 1992. **239**(4): p. 223-7.
14. Lechtenberg, R. and S. Gilman, *Speech disorders in cerebellar disease.* Ann Neurol, 1978. **3**(4): p. 285-90.
15. Vilis, T. and J. Hore, *Central neural mechanisms contributing to cerebellar tremor produced by limb perturbations.* J Neurophysiol, 1980. **43**(2): p. 279-91.
16. Manto, M., *Mechanisms of human cerebellar dysmetria: experimental evidence and current conceptual bases.* J Neuroeng Rehabil, 2009. **6**: p. 10.
17. Hore, J. and D. Flament, *Evidence that a disordered servo-like mechanism contributes to tremor in movements during cerebellar dysfunction.* J Neurophysiol, 1986. **56**(1): p. 123-36.
18. Paulson, H.L., *The spinocerebellar ataxias.* J Neuroophthalmol, 2009. **29**(3): p. 227-37.
19. Tankard, R.M., et al., *Detecting Expansions of Tandem Repeats in Cohorts Sequenced with Short-Read Sequencing Data.* Am J Hum Genet, 2018. **103**(6): p. 858-873.
20. Dolzhenko, E., et al., *Detection of long repeat expansions from PCR-free whole-genome sequence data.* Genome Res, 2017. **27**(11): p. 1895-1903.
21. Kobayashi, H., et al., *Expansion of intronic GGCCTG hexanucleotide repeat in NOP56 causes SCA36, a type of spinocerebellar ataxia accompanied by motor neuron involvement.* Am J Hum Genet, 2011. **89**(1): p. 121-30.
22. Bahlo, M. and C.J. Bromhead, *Generating linkage mapping files from Affymetrix SNP chip data.* Bioinformatics, 2009. **25**(15): p. 1961-2.
23. Arias, M., M. Garcia-Murias, and M.J. Sobrido, *Spinocerebellar ataxia 36 (SCA36): <<Costa da Morte ataxia>>.* Neurologia, 2017. **32**(6): p. 386-393.

24. Chao, M.J., et al., *Population-specific genetic modification of Huntington's disease in Venezuela.* PLoS Genet, 2018. **14**(5): p. e1007274.
25. Jayadev, S., et al., *Cambodian founder effect for spinocerebellar ataxia type 3 (Machado-Joseph disease).* J Neurol Sci, 2006. **250**(1-2): p. 110-3.
26. Valera, J.M., et al., *Prevalence of spinocerebellar ataxia 36 in a US population.* Neurol Genet, 2017. **3**(4): p. e174.
27. Coutelier, M., G. Stevanin, and A. Brice, *Genetic landscape remodelling in spinocerebellar ataxias: the influence of next-generation sequencing.* J Neurol, 2015. **262**(10): p. 2382-95.
28. Galatolo, D., et al., *Clinical application of next generation sequencing in hereditary spinocerebellar ataxia: increasing the diagnostic yield and broadening the ataxia-spasticity spectrum. A retrospective analysis.* Neurogenetics, 2018. **19**(1): p. 1-8.
29. Sullivan, R., et al., *Spinocerebellar ataxia: an update.* J Neurol, 2019. **266**(2): p. 533-544.