# *RFC1* Repeat Distribution in the Cypriot Population

## Study of a Large Cohort of Patients With Undiagnosed Ataxia and Non-Disease Controls

Christina Votsi, BSc, PhD, Marios Tomazou, BSc, MSc, PhD, Paschalis Nicolaou, BSc, PhD,
Marios C. Pantzaris, MD, Giorgos Pitsas, MD, Archontia Adamou, MD, Kleopas A. Kleopa, MD,
Eleni Zamba-Papanicolaou, MD, and Kyproula Christodoulou, BSc, MSc, PhD

**Correspondence**
Dr. Christodoulou
roula@cing.ac.cy

## Abstract

### Background and Objectives

The intronic biallelic AAGGG expansion in the replication factor C subunit 1 (*RFC1*) gene was recently associated with a phenotype combining cerebellar ataxia, neuropathy, and vestibular areflexia syndrome, as well as with late-onset ataxia. Following this discovery, studies in multiple populations extended the phenotypic and genotypic spectrum of this locus. Multiple benign and additional pathogenic configurations are currently known. Our main objectives were to study the prevalence of the pathogenic AAGGG expansion in the Cypriot population, to further characterize the *RFC1* repeat locus allele distribution, and to search for possible novel repeat configurations.

### Methods

Cypriot undiagnosed patients, in the majority presenting at least with cerebellar ataxia and non-neurologic disease controls, were included in this study. A combination of conventional methods was used, including standard PCR flanking the repeat region, repeat-primed PCR, long-range PCR, and Sanger sequencing. Bioinformatics analysis of already available in-house short-read whole-genome sequencing data was also performed.

### Results

A large group of undiagnosed patients (n = 194), mainly presenting with pure ataxia or with ataxia accompanied by neuropathy or additional symptoms, as well as a group of non-disease controls (n = 100), were investigated in the current study. Our findings include the diagnosis of 10 patients homozygous for the pathogenic AAGGG expansion and a high percentage of heterozygous AAGGG carriers in both groups. The benign $AAAAG_n$, $AAAGG_n$, and $AAGAG_n$ configurations were also identified in our cohorts. We also report and discuss the identification of 2 recently reported novel and possibly benign repeat configurations, $AAAGGG_n$ and $AAGAC_n$, thus confirming their existence in another distinct population, and we highlight an increased frequency of the $AAAGGG_n$ in the patient group, including a single case of homozygosity.

### Discussion

Our findings indicate the existence of genetic heterogeneity regarding the *RFC1* repeat configurations and that the AAGGG pathogenic expansion is a frequent cause of ataxia in the Cypriot population.

# Glossary

**ARCA** = autosomal recessive cerebellar ataxias; **CANVAS** = cerebellar ataxia with neuropathy and vestibular areflexia syndrome; **MSA** = multiple system atrophy; **RFC1** = replication factor C subunit 1; **WGS** = whole-genome sequencing.

# Introduction

Autosomal recessive cerebellar ataxias (ARCAs) are a clinically and genetically heterogeneous group of neurodegenerative diseases with an estimated prevalence of 3–6/100,000.[1] The most common pathologic feature of ARCAs is cerebellar degeneration, and therefore, progressive ataxia is the main symptom, usually accompanied by dysarthria, dysmetria, or oculomotor abnormalities. Other neurologic and extra-neurologic signs can also often be observed.[2] Recent methodological advances in molecular diagnosis over the last 2 decades have led to an expanded list of ARCA pathogenic variants in more than 100 genes,[1] including many novel genes. The replication factor C subunit 1 (RFC1) gene is one of the recently identified genes. A biallelic pentanucleotide (AAGGG) repeat expansion in RFC1 has been initially associated with cerebellar ataxia with neuropathy and vestibular areflexia syndrome (CANVAS) and other phenotypes of late-onset cerebellar ataxia[3,4] in European patients, with the number of repeats usually ranging from 400 to more than 2,000. Two other non-pathogenic polymorphic repeat configurations ($AAAAG_n$ and $AAAGG_n$) were also reported. Following initial observations, studies in different populations extended the heterogeneity of the locus by identifying additional pathogenic and non-pathogenic repeat configurations. These include the pathogenic $ACAGG_n$ motif identified in Asia Pacific cohorts[5] and the Maori-specific allele configuration of AAAGG(10-25)AAGGG(n) AAAGG(4-6) discovered within the Maori populations of New Zealand and the Cook Islands.[6] Likely non-pathogenic $AAGAG_n$ and $AGAGG_n$ motifs were described in a Canadian and Brazilian cohort.[2] The $AACGG_n$ motif was also described in a single non-CANVAS case in heterozygosity with another motif.[7] Very recently, a study of whole-genome sequencing (WGS) data of nearly 10,000 individuals supported pathogenicity for the already reported $AGAGG_n$ and non-pathogenicity for the $AAGAG_n$. This study also supported pathogenicity for 2 novel repeat expansions, the $AGGGC_n$ and $AAGGC_n$, as well as large uninterrupted or interrupted AAAGG expansions.[8] Another recent Australian population study described 3 additional most probably non-pathogenic repeats ($AGGGG_n$, $AAAGGG_n$, and $AAGAC_n$), by using the targeted long-read sequencing approach.[9]

The phenotypic spectrum of the RFC1 reported variants has been extended to include additional conditions to the initially described, such as pure cerebellar ataxia, isolated somatosensory impairment, and other parkinsonian disorders, including multiple system atrophy (MSA).[10] However, it is disputable if attributing a previously described disease, such as the CMT[11] or MSA,[12] to a novel gene/genetic mechanism is the correct approach. Alternatively, it has been suggested that phenotypic overlap is more likely to exist between diseases (i.e., CMT and

an RFC1-related disease with somatosensory loss; MSA and an RFC1-related disease with autonomic dysfunction and/or parkinsonism).[10] Furthermore, while the reported mean age at onset is high in most of the cases, patients with an earlier age at onset, such as the third decade or even earlier at 19,[13] 15,[14] or 10 years old,[15] have been described. It is also remarkable that smaller repeat expansion lengths (100–160 AAGGG repeats) have been reported as disease-causative for MSA.

Since identifying the RFC1 first pathogenic expansion, studies have been performed in control populations of different ethnicities and cohort sizes to estimate the carrier frequency ranging from 0.7% to 6.5%.[10] A more accurate estimation is expected to be performed in future analysis of large population data sets using improved bioinformatics tools. Many clinical cohorts with ataxia have also been studied, and biallelic pathogenic expansions have been identified in up to 38% of the tested patients in some of these cohorts. Therefore, the current data indicate that RFC1 expansions are likely the most frequent cause of recessive ataxias, and many patients remain undiagnosed.[10,16] More recently, the genotype spectrum of RFC1-related disease has been expanded by identifying truncating, frameshift, and nonsense variants in compound heterozygosity with an AAGGG expansion.[14,16,17] Regarding the disease pathophysiologic mechanism in the case of homozygous expansions, it proved to be puzzling and remains undetermined. Thus far, extensive functional studies have excluded any effect on RNA and protein expression or the function of DNA replication and repair machinery.[3] In contrast, in the compound heterozygous cases for a truncating variant and an expansion, a significant reduction of the mRNA was observed and suggested biallelic RFC1 conditional loss of function, thus shedding some light on the unexplained disease pathogenesis.[14,16,17]

Our group has studied families and sporadic cases of pure or complex ataxia phenotypes in the Cypriot population for more than 30 years. Thus far, the most frequent pathogenic variation in our population is the GAA repeat expansion in intron 1 of the FXN gene.[18,19] Additional pathogenic variants in ARCA and spastic ataxia genes have been identified in distinct families, some of which have been published.[20-23] However, a large pool of sporadic patients, excluded for the most common repeat expansions[24] and all the other variants identified in the population, remain genetically undiagnosed. Considering the appearance of the novel RFC1 gene as a frequent cause of adult-onset ataxia, we aimed to study the prevalence of the AAGGG repeat expansion and the other common RFC1 repeat locus alleles in a large group of undiagnosed patients presenting either with pure ataxia or with

ataxia accompanied by neuropathy or additional symptoms, in their majority at late onset. Few patients with an earlier age at onset were also included in this study group since such cases have been reported with *RFC1* pathogenic variants. A small number of clinically suspected CANVAS and neuropathy cases were also tested. In addition, a group of non-disease controls was studied to obtain a more reliable determination of the *RFC1* repeat locus alleles and their frequencies in the population. Furthermore, analysis of already available in-house short-read whole-genome sequencing data of non-ataxia Cypriot individuals was performed to investigate the presence of any other unknown repeat motif in our population. We hereby report our current findings, including the diagnosis of 10 homozygous patients for the pathogenic AAGGG expansion and the high percentage of heterozygous AAGGG carriers both in the group of patients and controls. We also report the identification of 2 of the recently reported novel benign repeat configurations $AAAGGG_n$ and $AAGAC_n$,[8,9] thus confirming their existence in another distinct population, and we highlight an increased frequency of the hexanucleotide $AAAGGG_n$ in the patient group.

## Methods

### Samples

One hundred eighty-seven Cypriot sporadic patients presenting at least with cerebellar ataxia were identified from the Neurogenetics Department (Cyprus Institute of Neurology and Genetics) database and included in the study. In their great majority (90%), they had adult-onset disease, and their current age was over 30. Any causes of acquired ataxia had been excluded in all of them. In their majority, they had also been previously excluded from the repeat expansions causing FRDA, SCA1, SCA2, SCA3, SC6, SCA7, SCA8, SCA10, SCA12, SCA17, and DRPLA and other recessive variants identified to cause pure or complicated ataxia in Cypriot families. Two available patients from a family with 3 affected individuals, the proband presenting full-blown CANVAS and 2 affected sisters presenting with cerebellar ataxia and sensory neuropathy (Figure 1), 3 additional sporadic patients with clinically suspected CANVAS, 2 sporadic patients with neuropathy, and a cohort of 100 non-disease Cypriot control individuals over 30 years old, were also included in the study.

### Standard Protocol Approvals, Registrations, and Patient Consents

This study has been approved by the National Bioethics Committee of Cyprus (ΕΕΒΚ/ΕΠ/2013/28). Written informed consent was obtained from all control individuals participating in the study. For the patients, no additional informed consent was requested as the test was in line with the original request for diagnostic testing.

### Flanking PCR and RP-PCR Analyses

To investigate the presence of an AAGGG pathogenic expansion, determine the configurations of any expansions, and genotype each individual, all patient and control DNA samples were analyzed following the PCR protocols previously described.[3] In brief, a standard PCR was performed for all the samples using a set of primers flanking the repeat region. The expected product corresponding to the reference allele $(AAAAG)_{11}$ should be 355 bp. A larger band indicates an intermediate expansion of any polymorphic repeat motifs. The absence of a product suggests a possible large expansion of the pathogenic repeat motif that could not be amplified under these conditions. All samples were also subjected to repeat-primed PCR (RP-PCR) analyses specific for the first described and most common AAAAG, AAAGG, and AAGGG configurations.[3] Additional RP-PCR analyses were performed for the samples excluded from any of these 3 motifs. These have been specific for the distinct reported configurations $ACAGG_n$, $AGAGG_n$, and $AAGAG_n$ and the 2 novel $AAAGGG_n$ and $AAGAC_n$ that we had identified in our population before they were published by other groups.[8,9] The current study uncovered these 2 repeats using different approaches, as described below in the results section. All RP-PCR products were analyzed on an ABI 3500xL Genetic Analyzer [Applied Biosystems (ABI), CA, USA], and the results were visualized using the GeneMapper v5 software (ABI).

---

**Figure 1** Pedigree of Family 1



The proband is indicated by the arrow.
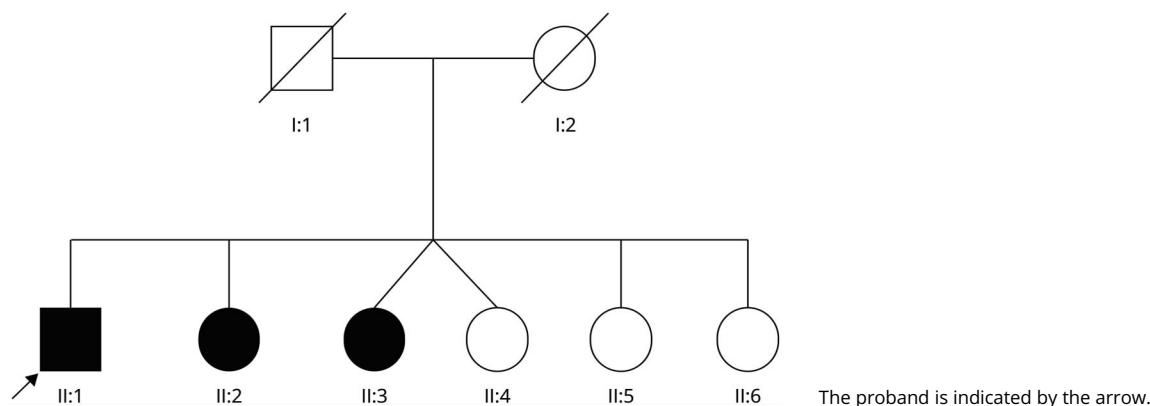
## Sanger Sequencing

Sanger sequencing was performed in the following cases: (1) To confirm the AAGGG motif in the patients with confirmed homozygous AAGGG expansion by the preliminary testing. Long-range PCR was used to amplify the repeat locus in the samples with no product in the preliminary flanking PCR and with positive RP-PCR for the AAGGG configuration. The Phusion polymerase was used in 30 μL volume following the reaction conditions previously described.[3] The PCR product was then used as the template for Sanger sequencing in both directions. (2) To investigate the samples resulting in peculiar RP-PCR patterns or gel electrophoresis bands through the preliminary screening. PCR products were excised and extracted from the gel using the NucleoSpin Gel and PCR clean-up kit (Macherey-Nagel) according to the manufacturer's instructions and were used as templates for Sanger sequencing in all cases. (3) To genotype the SNPs rs2066790, rs11096992, rs17584703, and rs6844176, which map inside the *RFC1* region. Primers amplifying the sequence encompassing each SNP were designed by us and are available on request. Standard PCR was performed, and the products were used as templates for Sanger sequencing.

Sanger sequencing was performed using the BigDye Terminator v1.1 Cycle Sequencing kit (ABI). Sequence traces were evaluated using the Sequence Analysis software (ABI). For SNP genotyping, traces were also automatically compared with the normal region sequences listed in the GenBank database using the SeqScape software (ABI).

## WGS Data Analysis

The aligned short-reads from WGS of a cohort of 96 Cypriot non-ataxia samples, available in-house, were visually inspected at the region containing the *RFC1* intron 2 using the Integrative Genomics Viewer software (IGV) (Web app v1.13.11 igv.js v 2.15.11, Broad Institute). The alignments were further analyzed using the ExpansionHunter De novo software (version 0.9.0)[4,25] to investigate for any other unknown repeat motif in our cohort.

## Allele Distribution Comparison

To compare the distribution of the *RFC1* alleles in the patient and control groups, we performed a $\chi^2$ test using the counts of the 7 conformations found in our population thus far (AAAAG$_{9-14}$, AAAAG$_n$, AAAGG$_n$, AAGGG$_n$, AAGAG$_n$, AAAGGG$_n$, and AAGAC$_n$) in a 2 × 7 contingency table. Additional pairwise comparisons across the different conformations, as well as the pathogenic AAGGG$_n$ and the likely non-pathogenic AAAGGG$_n$, vs all non-pathogenic patterns were performed.

## Data Availability

The data supporting this study's findings are available from the corresponding author on reasonable request.

## Results

We examined the prevalence of the *RFC1* expansions in the Cypriot population. We tested for the presence and nature of repeat expansions in 194 patients included in the study. Preliminary screening revealed no PCR product on flanking PCR in 10/194 (5.15%) individuals, thus indicating the presence of a biallelic expansion. Further investigation with RP-PCR revealed the presence of the AAGGG pathogenic expansion and the absence of AAAAG and AAAGG expansions. Subsequent long-range PCR enabled the amplification of large size products (corresponding to more than 500 repeats), and Sanger sequencing confirmed the presence of the AAGGG repeat sequence. Furthermore, these patients were genotyped at 4 informative SNPs defining a shared haplotype by most reported patients [rs2066790 (A/A), rs11096992 (A/A), rs17584703 (C/C), and rs6844176 (T/T)][3] and were also found to share this haplotype. Accurate sizing calculation was not possible due to an insufficient amount of DNA for Southern blotting. Long-read sequencing was also unavailable in our laboratory. The clinical features of these 10 patients are described in the subsection below and are also summarized in Table 1.

To complete the allelic characterization of the remaining individuals, we performed a preliminary genetic analysis with RP-PCR on known repeat configurations for all individuals presenting with at least one band <1,300 bp on flanking PCR, thus excluding homozygosity for the most common pathogenic AAGGG$_n$ or any other reported pathogenic expansions of more than ∼150 repeats (that could be detected with flanking PCR). Overall, the distribution of the distinct motif alleles showed differences between patients and controls (the allele counts and Chi-square calculation values are shown in eTables 1 and 2). The non-expanded AAAAG$_{9-14}$ allele was identified as the most common in the patient (45.7%) and control (56%) cohorts, followed by the expanded polymorphic AAAAG$_n$ allele (Table 2). The polymorphic expanded AAAGG$_n$ allele was identified in a low frequency (1.3%–2.0%). In contrast, a high frequency of the AAGGG$_n$ expanded pathogenic allele was observed in both groups, much higher and statistically significant in the patient's group (12.5%). Clinically, 2/27 patients heterozygous for the AAGGG expansion presented with full-blown CANVAS, and 7/27 presented with 2/3 typical symptoms (ataxia with sensory neuropathy [6/27] and ataxia with vestibulopathy [1/27]). For samples without fully determined allelic repeat combinations by analysis of the 3 common motifs, further analysis by RP-PCR on the other reported motifs revealed the presence of the AAGAG$_n$ allele in some individuals (7 patients and 3 controls). The ACAGG$_n$ and AGAGG$_n$ motifs were excluded from all tested samples. Sanger sequencing was employed to further investigate a selected number of samples presenting with uncharacteristic laddering RP-PCR patterns (such as atypical short saw-tooth patterns or patterns with gaps) that indicated the presence of complex alleles of more than one repeat motif. This investigation required gel excision

**Table 1** Clinical Features of the Patients Carrying the Homozygous AAGGG Pathogenic *RFC1* Repeat Expansion

| N | Age at onset | Age at examination | Fam/S | Symptom at onset | Disease progression | Cerebellar dysfunction | Peripheral neuropathy | Bilateral vestibular impairment | Autonomic dysfunction | Cough | MRI findings | Nerve conduction studies findings | Evoked potentials |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 60–65 | ~65 | Fam | Gait unsteadiness, slurred speech | Slow | Yes | Yes | Yes | No | No | White matter lesions, mild CA | Sensory axonal polyneuropathy | Normal VEPs, abnormal SSEPs, and BAEPs |
| 2 | 60–65 | ~70 | Fam | Gait unsteadiness | Slow | Yes | Yes | No | No | No | White matter lesions | Sensory axonal polyneuropathy | n/a |
| 3 | 50–55 | ~65 | S | Gait unsteadiness | Slow | Yes | Yes | Yes | Yes (postural hypotension, sweating issues, easy abdominal bloating, xerophthalmia, occasional urinary complaints, syncope episodes) | Yes | CA | Sensory and mild motor axonal polyneuropathy | n/a |
| 4 | 60–65 | ~65 | S | Gait unsteadiness | Slow | Yes | Yes | No | No | No | CA | Sensory axonal polyneuropathy | Abnormal VEPs and SSEPs |
| 5 | 65–70 | ~70 | S | Neurogenic cough, loss of consciousness[a] | Slow | No | Yes | Yes | Yes (constipation, urinary incontinence and impotence) | Yes | No major abnormalities | Sensory axonal polyneuropathy | n/a |
| 6 | 50–55 | ~60 | S | Gait unsteadiness and cough | Slow | Yes | Yes | No | No | Yes | Mild CA | Sensory axonal polyneuropathy | n/a |
| 7 | 55–60 | ~60 | S | Gait unsteadiness with occasional numbness of LL | Slow | Yes | No[b] | No | Yes (bowl disturbances and diarrhea episodes) | Yes | Mild CA | n/a | n/a |
| 8 | 60–65 | ~65 | S | Gait unsteadiness | Slow | Yes | Yes | No | No | No | CA | Sensory axonal polyneuropathy | n/a |
| 9 | 40–45 | ~55 | S | Gait unsteadiness | Slow | Yes | Yes | Yes | Yes (urinary incontinence) | No | CA | Sensory axonal polyneuropathy | Normal VEPs and BAEPs |
| 10 | 45–50 | ~55 | S | Gait unsteadiness, long standing pain below left knee | Slow | Yes | Yes | Yes | No | No | CA, brainstem and spinal cord atrophy | Sensory axonal polyneuropathy | Normal VEPs |

Abbreviations: BAEPs = brainstem auditory evoked potentials; CA = cerebellar atrophy; LL = lower limbs; SSEPS = somatosensory evoked potentials; VEPs = visual evoked potentials.
[a] Loss of consciousness was triggered by dysautonomia.
[b] NCS had not been performed for this patient. Clinical signs of pallesthesia disturbance were observed peripherally at both legs, which is not a definite sign of peripheral neuropathy (this could be due to type 2 diabetes complications).

**Table 2** Allele Frequencies (%) of the *RFC1* Repeat Identified Motifs in the Cypriot Cases and Control Cohorts

| Repeat motif | Patients including AAGGGn[a] homozygotes | Patients excluding AAGGGn[a] homozygotes | Controls |
|---|---|---|---|
| AAAAG$_{(9-14)}$ | 45.7 | 48.3 | 56 |
| AAAAGn[a] | 32.2 | 34 | 31.5 |
| AAAGGn[a] | 1.3 | 1.4 | 2.0 |
| AAGAGn[a] | 1.9 | 2.0 | 1.5 |
| AAGGGn[a] | 12.5 | 7.6 | 3.0 |
| AAAGGGn[a] | 6.4 | 6.8 | 3.5 |
| AAGACn[a] | — | — | 2.5 |

[a] n represents expanded repeats.

and extraction of flanking PCR products before performing Sanger sequencing. This analysis revealed the existence of an allele of ~60–70 repeats, consisting mainly of the hexanucleotide repeat motif AAAGGG (~48 repeats), combined with a small segment of AAGGG repeats (in the majority of samples) or AAAGG repeats (eFigure 1). RP-PCR on this motif was also performed and confirmed the presence of a decremental saw-tooth pattern in all samples. This pattern appeared with a delay compared with the primer position, thus indicating the presence of different sequence motifs preceding the repeat. In these cases, the initial RP-PCR obtained atypical short saw-tooth patterns on the AAGGG or the AAAGG motif, corresponding to the repeats neighboring the AAAGGG motif (Figure 2). The presence of this repeat was tested by RP-PCR in all samples presenting with the same atypical RP-PCR during the preliminary analysis and confirmed its presence. The frequency of this repeat was also tested in the cohort of the non-disease control samples. Of interest, it was higher in the patients' group (6.4%) than the controls (3.5%); however, non-significant, and a single homozygote with slowly progressive ataxia by the age of 54 and axonal polyneuropathy was identified. Regarding the remaining AAAGGG expansion heterozygote patients, 3/23 also carried the AAGGG expansion; one presented with ataxia and neuropathy. None of the remaining AAAGGG heterozygote patients presented with neuropathy or

**Figure 2** Representative RP-PCR Plots of Cases Presenting With Atypical Patterns by the Preliminary Analysis of Motifs AAAAG, AAAGG, and AAGGG and Confirmed With an Allele Consisted Mainly of the AAAGGG Motif
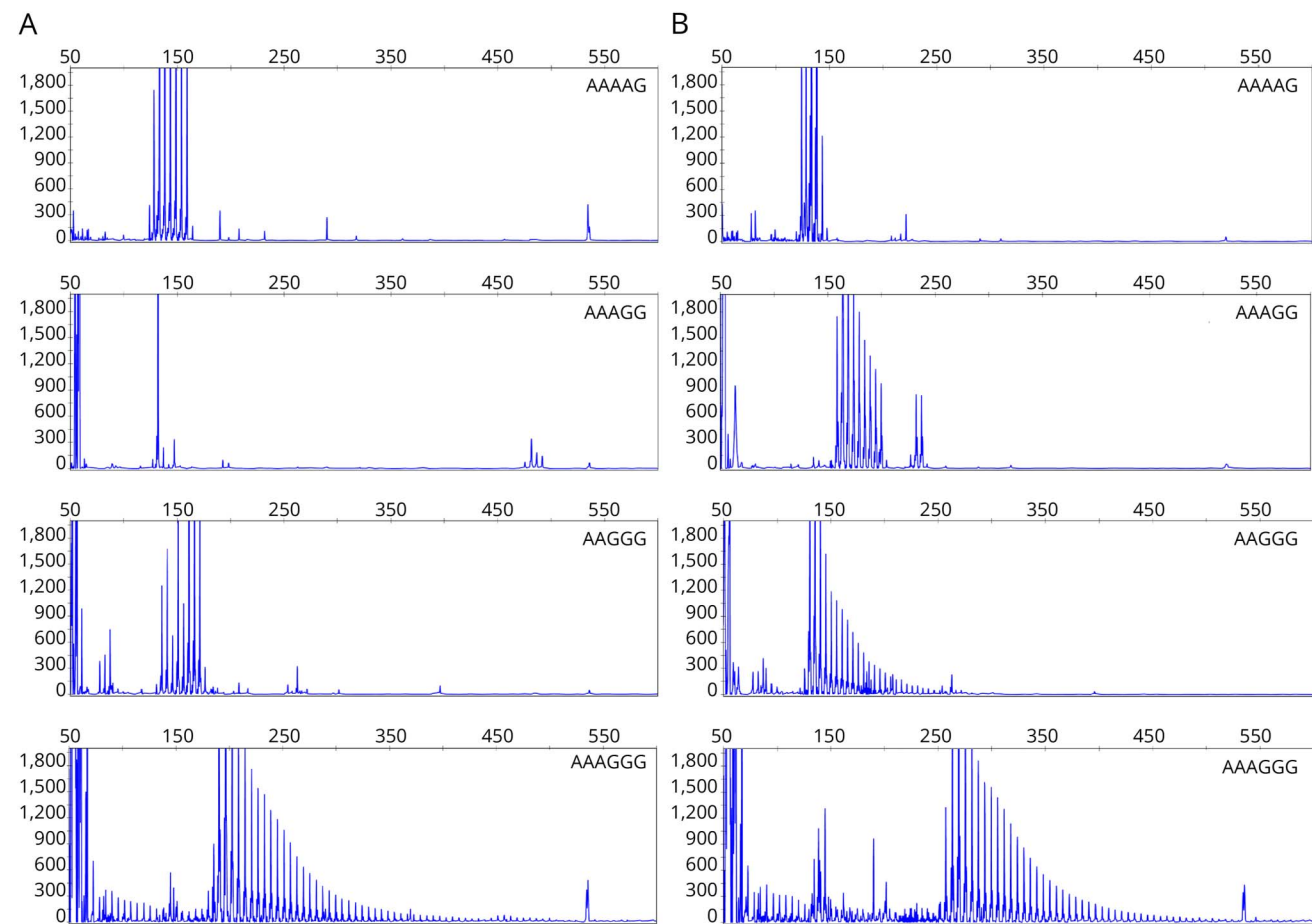


(A) Plots of an individual carrying the normal AAAAG$_{11}$ allele and an expanded allele mainly of AAAGGG repeats, a small segment of AAGGG and other repeats as indicated by Sanger sequencing. (B) Plots of an individual carrying the expanded AAGGG allele and an expanded allele mainly of AAAGGG repeats and a smaller segment of AAAGG repeats.

**Figure 3** IGV Snapshots of the *RFC1* Repeat Region Obtained by the Analysis of Whole-Genome Short-Read Sequencing Data
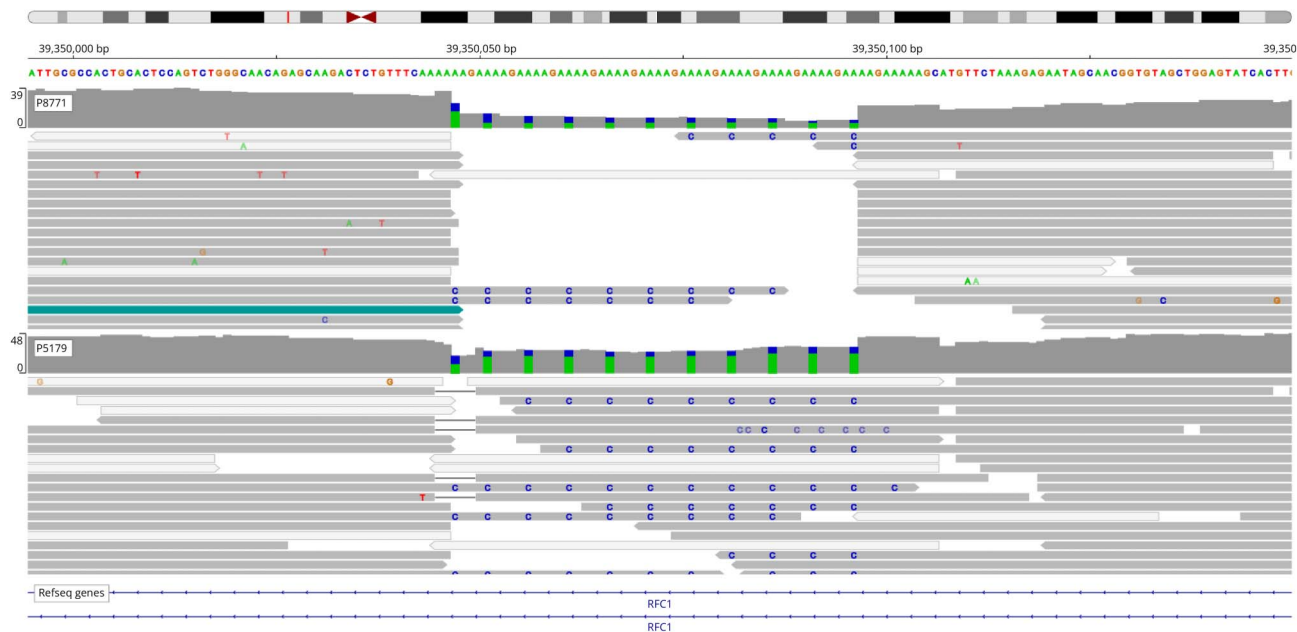


The data of 2 Cypriot individuals from a CMT1A family (not presenting ataxia features), with the AAGAC repeat expansion, are displayed. The first individual was found to carry one allele with the pathogenic AAGGG expansion and one allele with the expanded motif AAGAC. The second individual was found to carry one allele with the non-expanded AAAAG repeat and one allele with the expanded motif AAGAC. These data were further confirmed by the use of ExpansionHunter De novo software, as well as experimentally by flanking PCR and RP-PCR targeting this novel and the other common repeat motifs. IGV = Integrative Genomics Viewer.

vestibulopathy in addition to ataxia. For 6 out of 184 non-$AAGGG_n$ homozygous patients studied, the above analyses were insufficient to resolve the observed complexity in RP-PCR patterns and conclude on the second allele synthesis. Therefore, they have been excluded from the allele frequency estimation that is currently presented (Table 2).

In addition to the above, we identified another repeat motif through WGS data analysis. This expanded $AAGAC_n$ repeat was found only in compound heterozygosity with the expanded $AAGGG_n$ repeat or the expanded $AAAAG_n$ repeat in this family (Figure 3). In contrast to the other polymorphic repeats, the allele carrying this repeat was not amplifiable with flanking PCR, thus indicating a high number of repeats similar to the pathogenic $AAGGG_n$ repeat. However, accurate sizing of this expansion was impossible due to an insufficient amount of DNA for Southern blot experiments. RP-PCR analysis confirmed the presence of a decremental saw-tooth pattern in all candidate family members, as indicated by WGS data. Further screening by RP-PCR to estimate its frequency in the control population enabled its identification in 5/100 individuals in compound heterozygosity with the expanded AAAAG, or the non-expanded AAAAG, or the polymorphic expanded AAGAG repeats.

### Clinical Features of the Patients Confirmed With Biallelic *RFC1* Expansion

The clinical features of the patients confirmed with homozygous AAGGG expansions (6 males and 4 females) were retrospectively collected and are summarized in Table 1. Most patients (9/10) presented with adult onset gait unsteadiness as their first symptom and shared cerebellar dysfunction and peripheral (mostly sensory axonal) neuropathy features. Cerebellar dysfunction was absent only from patient number 5. This patient had neurogenic cough and loss of consciousness as the first symptoms. The progression of the disease was slow in all patients. Bilateral vestibular impairment, autonomic dysfunction, and chronic cough, common in CANVAS cases, were reported for some patients (5/10, 4/10, and 4/10, respectively). Cerebellar atrophy, either mild or moderate, was detected in most cases.

Only 4 of them, one of which was the family 1 proband, presented with the 3 symptoms of full-blown CANVAS, cerebellar ataxia, neuropathy, and vestibulopathy. In 4 of the remaining cases with no vestibulopathy, one of which was the second member of family 1, cerebellar ataxia and neuropathy coexisted. In contrast, cerebellar ataxia was absent from a single case of neuropathy and vestibulopathy. A single case also presented with pure ataxia without neuropathy.

## Discussion

We present a study investigating the *RFC1* gene polymorphic and pathogenic expanded repeat configurations in the Cypriot population. A large group of genetically undiagnosed patients (194) consisted mainly of adult and a few younger onset ataxia

cases, along with a few CANVAS suspected and neuropathy cases, and a group of non-disease control individuals have been included in this investigation. Genetically undiagnosed younger onset cases were also included because such cases have been reported as positive for *RFC1* expansions.[13-15] In addition, the growing number of studies performed thus far on this relatively recently described disease indicated an expansion of the phenotypic spectrum and the genetic heterogeneity, which possibly remain not fully elucidated.[10,12,13,15,26-30] Therefore, we cannot exclude the possibility of identifying positive younger-age cases. Ten out of 194 patients tested (5.2%) have been diagnosed with the pathogenic AAGGG repeat expansion in a homozygous state. More specifically, it was found in 4 out of 7 cases (57%) presenting with full-blown CANVAS, in one out of 2 cases with neuropathy (50%), and in 5 out of 185 cases (2.7%) presenting at least with ataxia as one of their prominent symptoms. Overall, the observed phenotypes of the diagnosed patients support the broad clinical spectrum associated with *RFC1* pathogenic expansions.[10] Similar to other reports, all diagnosed patients had at least 2 of the 3 typical features of CANVAS (mostly CA and sensory neuropathy) except for one presenting with pure ataxia. Full-blown CANVAS was presented only in 4 out of 10 cases. Moreover, a retrospective collection of clinical and other personal information of the diagnosed cases indicated that 4 out of the 10 cases originate from the same village, thus indicating the possible existence of a founder effect.

The currently reported diagnostic yield for the biallelic AAGGG repeat expansion in the group of CANVAS cases (57%) falls within the range reported by other studies (36%–92%).[3,4,9,13,26,31] Regarding our main cohort of undiagnosed cases presenting with pure or complex ataxia phenotypes, the calculated diagnostic yield for a homozygous expansion is much lower (2.7%). This percentage is close to the lower range of the other reported yields in different cohorts of mixed pure and complex ataxia cases rather than a full CANVAS phenotype (0.6% Canadian,[2] 4.3% Brazilian,[2] 1.0% French,[31] 1.6% Dutch,[32] 1.6% Greek,[33] 7.6% German,[26] 14% Caucasian,[13] and 15.9% British[3]). A more detailed clinical description of all our patients or exclusion of the young onset cases (n = 17) could slightly increase this percentage. Moreover, we report a higher percentage of patients, compared with the controls, that have been heterozygous for the AAGGG allele, including a case with a young age at onset, 2 typical CANVAS cases, and 7 cases presenting with 2/3 typical symptoms, thus indicating the possible presence of a second intragenic variant that remains to be determined.

The estimated frequency of the detected alleles in the non-disease control chromosomes revealed $AAAAG_{(9-14)}$ as the most frequent allele (57%) followed by $AAAAG_n$ (33%), as in the majority of the reported populations ($AAAAG_{11}$, 75%–85%; $AAAAG_n$, 7%–13%).[2,3] However, compared with other studies, the frequency of both alleles is quite different (lower for $AAAAG_{9-14}$ and higher for $AAAAG_n$). $AAAGG_n$ is very rare (2%). In contrast, the frequency of the pathogenic $AAGGG_n$ is relatively high in the control group (3%) and much higher in the patient group (12.5%). Even if excluding the 10 $AAGGG_n$

homozygous cases, the percentage (7.6%) is still much higher compared with the controls, thus suggesting that intragenic *RFC1* variants might also exist in some of these patients in compound heterozygosity with the expanded $AAGGG_n$ allele, pending identification. In addition to the 3 common conformations tested first, the polymorphic $AAGAG_n$ was identified at 1.9% and 1.5% in the patient and control groups, respectively. We also report the identification of the hexanucleotide AAAGGG repeat expansion in both cohorts. In the course of the final stages of our study, this novel repeat expansion has also been reported by 2 other groups. Dominik et al. (2023) identified this and other novel repeats by performing bioinformatics analysis on WGS data of 10,000 individuals from the 100,000 Genomes Project, and they categorized it as likely non-pathogenic. Scriba et al. (2023) identified this hexanucleotide repeat by using targeted long-read sequencing in 2 patients of an Australian cohort screened for *RFC1* expansions, and they did not make a clear conclusion about its pathogenicity. Further screening using RP-PCR resulted in its identification in 3 additional cases, thus concluding to a total number of 5 cases out of 242 (2.1%), all compound heterozygotes. In the current study, conventional methods were used to investigate further cases resulting in uncharacteristic laddering RP-PCR patterns on the first tested repeats and enabled the discovery of this hexanucleotide repeat. As there was a detectable flanking PCR product on gel electrophoresis in all candidate cases (ranging from ~700 to ~1,000 bp) that was distinct from the reference allele size, Sanger sequencing was initially used and successfully uncovered the $AAAGGG_n$ motif, which was subsequently also confirmed by RP-PCR. This repeat has been reported only in compound heterozygosity with the AAGGG or other non-pathogenic expansions.[8,9] Homozygosity in a single patient presenting with ataxia and neuropathy is reported for the first time in the current study. Furthermore, compared with Dominik et al. (2023), who found a similar frequency of the compound heterozygotes AAGGG/AAAGGG in patients (0.6%) and controls (0.4%), we identified 3 compound heterozygous AAGGG/AAAGGG patients (1.6%) and none in the controls (0%). However, we report an overall comparatively higher frequency of this repeat allele in the patient's group (6.4%) compared with the controls (3.5%), which resembles the difference in frequency we observed for the pathogenic $AAGGG_n$ allele. Therefore, we still cannot conclude on its pathogenicity. In addition, as already discussed by Scriba et al. (2023), the size of this expansion reflects more the size of the other benign configurations. Still, the GC content (50%) is closer to that of the pathogenic configurations (60%–80%) than the benign (20%–40%), thus suggesting that pathogenicity depends more on repeat configuration than the expansion size.

We also report identifying the repeat configuration $AAGAC_n$ in heterozygosity with the pathogenic $AAGGG_n$ repeats or other non-pathogenic configurations, with the limitation of not detecting its size. This novel configuration was also reported in the course of the final stages of our study by Scriba et al. (2023), who used targeted long-read sequencing to identify this repeat, similar to the identification of the $AAAGGG_n$. This approach enabled the size determination at ~270 repeats, and the

identified frequency in this cohort was 1.65% (4/242 patients). Pathogenicity was not supported clearly. We identified this repeat by analyzing short-read WGS data. This approach could not define the expansion size. RP-PCR further confirmed the presence of this configuration, and flanking PCR indicated a possible large size since there was no amplifiable product. Our hypothesis was supported by the size reported in the Australian patients. Thus far, this novel repeat allele was identified in our control cohort only, and its estimated carrier frequency was 5%. In addition, in the CMT1A family initially identified, it was found in compound heterozygosity with the pathogenic AAGGG expansion both in one CMT1A patient and her unaffected mother. Therefore, based on these observations, it is likely a non-pathogenic configuration.

Future investigation is expected to shed more light on the effect of these novel repeat motifs and the puzzling disease pathogenetic mechanism in general. Thus far, structure functional analyses have been performed in vitro to study the formation of possible unusual DNA structures by short pathogenic AAGGG repeats compared with the non-pathogenic AAAAG repeats.[34,35] Of interest, it was shown that the pathogenic repeats can form triplex and quadruplex structures, thus leading to replication stalling in an orientation-dependent manner.[34] Moreover, it was shown that a G-quadruplex ligand (TMPyP4) binds specifically to the pathogenic repeats' quadruplexes.[35] These findings are promising toward understanding the disease pathogenicity and development of targeted therapies.

In conclusion, we hereby present our study of *RFC1* polymorphic and pathogenic configurations in the Cypriot population, which resulted in interesting findings. We report the confirmation of a homozygous $AAGGG_n$ pathogenic expansion in one neuropathy and 9 adult-onset ataxia cases presenting either with full-blown or incomplete CANVAS clinical symptoms. Of interest, 4 of these patients were found to have a common village origin. We also report the identification of a high percentage of heterozygosity for the pathogenic $AAGGG_n$ expanded allele in the remaining tested non-homozygous patients, thus indicating the possible existence of intragenic variants which are pending identification. In addition, we confirm the existence of the recently reported $AAAGGG_n$ and $AAGAC_n$ repeat motifs in another population and we describe their discovery in our cohorts using distinct approaches from those reported. We also highlight an increased frequency of the alleles containing the $AAAGGG_n$ repeat in the patient group compared with the controls, and we report a single case of homozygosity for this repeat motif for the first time. In addition, we report the existence of 6 cases with unusual RP-PCR patterns on the tested motifs that did not enable us to conclude on the genotypes, thus indicating the possible existence of additional unknown configurations that will expand more the genetic heterogeneity of the *RFC1* locus. Overall, our findings (including a high carrier frequency of the expansion) indicate that the *RFC1* pathogenic expansion is a frequent cause of ataxia in the Cypriot population, and therefore, prioritizing the *RFC1* testing on the patients fulfilling the CANVAS symptoms or at least 2 of the main key symptoms is significant.

## Appendix Authors

| Name | Location | Contribution |
|---|---|---|
| Christina Votsi, BSc, PhD | Neurogenetics Department, The Cyprus Institute of Neurology and Genetics | Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design; analysis or interpretation of data |
| Marios Tomazou, BSc, MSc, PhD | Bioinformatics Department, The Cyprus Institute of Neurology and Genetics | Drafting/revision of the manuscript for content, including medical writing for content; analysis or interpretation of data |
| Paschalis Nicolaou, BSc, PhD | Neurogenetics Department, The Cyprus Institute of Neurology and Genetics | Drafting/revision of the manuscript for content, including medical writing for content; analysis or interpretation of data |
| Marios C. Pantzaris, MD | Neuroimmunology Department, The Cyprus Institute of Neurology and Genetics | Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data |
| Giorgos Pitsas, MD | Neuroimmunology Department, The Cyprus Institute of Neurology and Genetics | Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data |
| Archontia Adamou, MD | Neuroepidemiology Department, The Cyprus Institute of Neurology and Genetics | Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data |
| Kleopas A. Kleopa, MD | Neuroscience Department and Center for Neuromuscular Disorders, The Cyprus Institute of Neurology and Genetics | Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data |
| Eleni Zamba-Papanicolaou, MD | Neuroepidemiology Department and Center for Neuromuscular Disorders, The Cyprus Institute of Neurology and Genetics | Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data |
| Kyproula Christodoulou, BSc, MSc, PhD | Neurogenetics Department, The Cyprus Institute of Neurology and Genetics | Drafting/revision of the manuscript for content, including medical writing for content; study concept or design |

# References

1. Synofzik M, Nemeth AH. Recessive ataxias. *Handb Clin Neurol.* 2018;155:73-89. doi:10.1016/B978-0-444-64189-2.00005-6

2. Akcimen F, Ross JP, Bourassa CV, et al. Investigation of the RFC1 repeat expansion in a Canadian and a Brazilian ataxia cohort: identification of novel conformations. *Front Genet.* 2019;10:1219. doi:10.3389/fgene.2019.01219

3. Cortese A, Simone R, Sullivan R, et al. Biallelic expansion of an intronic repeat in RFC1 is a common cause of late-onset ataxia. *Nat Genet.* 2019;51(4):649-658. doi:10.1038/s41588-019-0372-4

4. Rafehi H, Szmulewicz DJ, Bennett MF, 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):151-165. doi:10.1016/j.ajhg.2019.05.016

5. Scriba CK, Beecroft SJ, Clayton JS, et al. A novel RFC1 repeat motif (ACAGG) in two Asia-Pacific CANVAS families. *Brain.* 2020;143(10):2904-2910. doi:10.1093/brain/awaa263

6. Beecroft SJ, Cortese A, Sullivan R, et al. A Maori specific RFC1 pathogenic repeat configuration in CANVAS, likely due to a founder allele. *Brain.* 2020;143(9):2673-2680. doi:10.1093/brain/awaa203

7. Abramzon Y, Dewan R, Cortese A, et al. Investigating RFC1 expansions in sporadic amyotrophic lateral sclerosis. *J Neurol Sci.* 2021;430:118061. doi:10.1016/j.jns.2021.118061

8. Dominik N, Magri S, Curro R, et al. Normal and pathogenic variation of RFC1 repeat expansions: implications for clinical diagnosis. *Brain.* 2023;146(12):5060-5069. doi:10.1093/brain/awad240

9. Scriba CK, Stevanovski I, Chintalaphani SR, et al. RFC1 in an Australasian neurological disease cohort: extending the genetic heterogeneity and implications for diagnostics. *Brain Commun.* 2023;5(4):fcad208. doi:10.1093/braincomms/fcad208

10. Davies K, Szmulewicz DJ, Corben LA, Delatycki M, Lockhart PJ. RFC1-related disease: molecular and clinical insights. *Neurol Genet.* 2022;8(5):e200016. doi:10.1212/NXG.0000000000200016

11. Lipponen J, Helisalmi S, Raivo J, et al. Molecular epidemiology of hereditary ataxia in Finland. *BMC Neurol.* 2021;21(1):382. doi:10.1186/s12883-021-02409-z

12. Wan L, Chen Z, Wan N, et al. Biallelic intronic AAGGG expansion of RFC1 is related to multiple system atrophy. *Ann Neurol.* 2020;88(6):1132-1143. doi:10.1002/ana.25902

13. Cortese A, Tozza S, Yau WY, et al. Cerebellar ataxia, neuropathy, vestibular areflexia syndrome due to RFC1 repeat expansion. *Brain.* 2020;143(2):480-490. doi:10.1093/brain/awz418

14. Ronco R, Perini C, Curro R, et al. Truncating variants in RFC1 in cerebellar ataxia, neuropathy, and vestibular areflexia syndrome. *Neurology.* 2023;100(5):e543-e554. doi:10.1212/WNL.0000000000201486

15. Ando M, Higuchi Y, Yuan JH, et al. Genetic and clinical features of cerebellar ataxia with RFC1 biallelic repeat expansions in Japan. *Front Neurol.* 2022;13:952493. doi:10.3389/fneur.2022.952493

16. Arteche-Lopez A, Avila-Fernandez A, Damian A, et al. New Cerebellar Ataxia, Neuropathy, Vestibular Areflexia Syndrome cases are caused by the presence of a nonsense variant in compound heterozygosity with the pathogenic repeat expansion in the RFC1 gene. *Clin Genet.* 2023;103(2):236-241. doi:10.1111/cge.14249

17. Benkirane M, Da Cunha D, Marelli C, et al. RFC1 nonsense and frameshift variants cause CANVAS: clues for an unsolved pathophysiology. *Brain.* 2022;145(11):3770-3775. doi:10.1093/brain/awac280

18. Zamba-Papanicolaou E, Koutsou P, Daiou C, Gaglia E, Georghiou A, Christodoulou K. High frequency of Friedreich's ataxia carriers in the Paphos district of Cyprus. *Acta Myol.* 2009;28(1):24-26.

19. Dean G, Chamberlain S, Middleton L. Friedreich's ataxia in Kathikas-Arodhes, Cyprus. *Lancet.* 1988;1(8585):587. doi:10.1016/s0140-6736(88)91378-5

20. Nicolaou P, Georghiou A, Votsi C, Middleton LT, Zamba-Papanicolaou E, Christodoulou K. A novel c.5308_5311delGAGA mutation in Senataxin in a Cypriot family with an autosomal recessive cerebellar ataxia. *BMC Med Genet.* 2008;9:28. doi:10.1186/1471-2350-9-28

21. Votsi C, Zamba-Papanicolaou E, Middleton LT, Pantzaris M, Christodoulou K. A novel GBA2 gene missense mutation in spastic ataxia. *Ann Hum Genet.* 2014;78(1):13-22. doi:10.1111/ahg.12045

22. Nicolaou P, Tanteles GA, Votsi C, et al. A novel CLN6 variant associated with juvenile neuronal ceroid lipofuscinosis in patients with absence of visual loss as a presenting feature. *Front Genet.* 2021;12:746101. doi:10.3389/fgene.2021.746101

23. Votsi C, Ververis A, Nicolaou P, Christou YP, Christodoulou K, Zamba-Papanicolaou E. A novel SPG7 gene pathogenic variant in a cypriot family with autosomal recessive spastic ataxia. *Front Genet.* 2021;12:812640. doi:10.3389/fgene.2021.812640

24. Votsi C, Zamba-Papanicolaou E, Georghiou A, et al. Investigation of SCA10 in the cypriot population: further exclusion of SCA dynamic repeat mutations. *J Neurol Sci.* 2012;323(1-2):154-157. doi:10.1016/j.jns.2012.09.006

25. Dolzhenko E, van Vugt J, Shaw RJ, et al. Detection of long repeat expansions from PCR-free whole-genome sequence data. *Genome Res.* 2017;27(11):1895-1903. doi:10.1101/gr.225672.117

26. Gisatulin M, Dobricic V, Zuhlke C, et al. Clinical spectrum of the pentanucleotide repeat expansion in the RFC1 gene in ataxia syndromes. *Neurology.* 2020;95(21):e2912-e2923. doi:10.1212/WNL.0000000000010744

27. Traschutz A, Cortese A, Reich S, et al. Natural history, phenotypic spectrum, and discriminative features of multisystemic RFC1 disease. *Neurology.* 2021;96(9):e1369-e1382. doi:10.1212/WNL.0000000000011528

28. Curro R, Salvalaggio A, Tozza S, et al. RFC1 expansions are a common cause of idiopathic sensory neuropathy. *Brain.* 2021;144(5):1542-1550. doi:10.1093/brain/awab072

29. Tagliapietra M, Cardellini D, Ferrarini M, et al. RFC1 AAGGG repeat expansion masquerading as chronic idiopathic axonal polyneuropathy. *J Neurol.* 2021;268(11):4280-4290. doi:10.1007/s00415-021-10552-3

30. Reyes-Leiva D, Aldecoa I, Gelpi E, Rojas-Garcia R. Motor neuron involvement expands the neuropathological phenotype of late-onset ataxia in RFC1 mutation (CANVAS). *Brain Pathol.* 2022;32(4):e13051. doi:10.1111/bpa.13051

31. Montaut S, Diedhiou N, Fahrer P, et al. Biallelic RFC1-expansion in a French multicentric sporadic ataxia cohort. *J Neurol.* 2021;268(9):3337-3343. doi:10.1007/s00415-021-10499-5

32. Ghorbani F, de Boer-Bergsma J, Verschuuren-Bemelmans CC, et al. Prevalence of intronic repeat expansions in RFC1 in Dutch patients with CANVAS and adult-onset ataxia. *J Neurol.* 2022;269(11):6086-6093. doi:10.1007/s00415-022-11275-9

33. Kontogeorgiou Z, Kartanou C, Tsirligkani C, et al. Biallelic RFC1 pentanucleotide repeat expansions in Greek patients with late-onset ataxia. *Clin Genet.* 2021;100(1):90-94. doi:10.1111/cge.13960

34. Hisey JA, Radchenko EA, Ceschi S, et al. Pathogenic CANVAS (AAGGG)(n) repeats stall DNA replication due to the formation of alternative DNA structures. *Nucleic Acids Res.* 2024;gkae124. doi:10.1093/nar/gkae124

35. Abdi MH, Zamiri B, Pazuki G, Sardari S, Pearson CE. Pathogenic CANVAS-causing but not nonpathogenic RFC1 DNA/RNA repeat motifs form quadruplex or triplex structures. *J Biol Chem.* 2023;299(10):105202. doi:10.1016/j.jbc.2023.105202