Spinocerebellar ataxia type 10 (SCA10): Mutation analysis and common haplotype based inference suggest its rarity in Indian population

Divya Goel[a,b], Varun Suroliya[a], Uzma Shamim[a], Aradhna Mathur[a], Mohammed Faruq[a,*]

[a] Genomics and Molecular Medicine, CSIR-Institute of Genomics and Integrative Biology (CSIR -IGIB), Mall Road, Delhi, 110007, India
[b] National Institute of Pharmaceutical Education and Research, Guwahati, C/O NITES Institute of Technology and Science, NH-37, Shantipur, Mirza, Assam, 781125, India

A B S T R A C T

Spinocerebellar ataxia type 10 (SCA10) is a rare autosomal dominant cerebellar ataxia caused by nucleotide ATTCT expansion in *ATXN10* gene. SCA10 has been reported in patients of cerebellar ataxia from Amerindian/ Latin America and in East Asian ancestry. A common founder has been ascribed to the origin of ATTCT repeat expansion mutation in both the population. Here we present our investigation of the SCA10 pentanucleotide repeat expansion in **461** SCA patients of the Indian population. The analysis of multi-ethnic at-risk haplotype C-(ATTCT)n-GGC was performed using genotype data of various ethnic population included in the 1000 Genomes Project (KGP) to infer the prevalence of at-risk haplotype in the Indian populations. Unsurprisingly, none of the patient's DNA samples with (ATTCT)n expansion was observed in pathological range, however, the observed normal range of (ATTCT)n was 8–22 repeats, suggesting very rare or absence of the occurrence of SCA10 in Indian SCA patients. The at-risk haplotype, CGGC was found to be the most prevalent haplotype across different populations and no segregation of CGGC haplotype with large normal or small normal ATTCT repeats length was observed. However, on extended haplotype analysis, some lineage of CGGC with a flanking divergence at 5′ end was observed specifically in the American or East Asian population but not in other population in KGP dataset. Together, these evidence points towards the absence of SCA10 in Indian population and haplotype-based analysis also suggests its occurrence to be rare in South Asian, European and African population. Further investigations are required to establish the present finding.

Significance: The implications of the findings of this study are 1.) For the diagnostic work-up of SCAs in the Indian population and to decide upon inclusion of SCA10 in panel based genetic investigations even for Indians living abroad. 2.) The haplotype based inference of its presumptive prevalence through the estimation of at-risk haplotype using population genetics approach (South-Asians as the background) allowed us to estimate the possible absence of SCA10 in Indian population. SCA10 is a rare autosomal dominant cerebellar ataxia mostly reported among SCA patients from Latin America and recently described in East Asia population. The genetic study of SCA10 performed in the unrelated Indian spinocerebellar ataxia patients with heterogeneous ethnicity confirmed its absence from the Indian population and that conforms to population genetic based inference of its rarity or absence. 3.) This approach may be adopted for the screening of other subtypes of SCAs, i.e. other rare SCAs e.g. SCA31, SCA36, and SCA37.

1. Introduction

Spinocerebellar ataxia type 10 (SCA10) is a rare autosomal dominant cerebellar ataxia caused by pentanucleotide (ATTCT) repeat expansion in intron 9 of the *ATXN10* gene which maps to chr22q13.3 region [1–3]. The number of (ATTCT)n repeats commonly found among different populations ranged from 10 to 29, whereas the pathological expanded alleles ranged from 800 to 4500 repeats in affected patients. Furthermore, the intermediate size alleles of repeats 280 to 850 exhibited reduced penetrance [2,4,5]. SCA10 is characterized by progressive cerebellar dysfunctions together with epilepsy and peripheral neuropathy. The first identification of SCA10 in two Mexican families
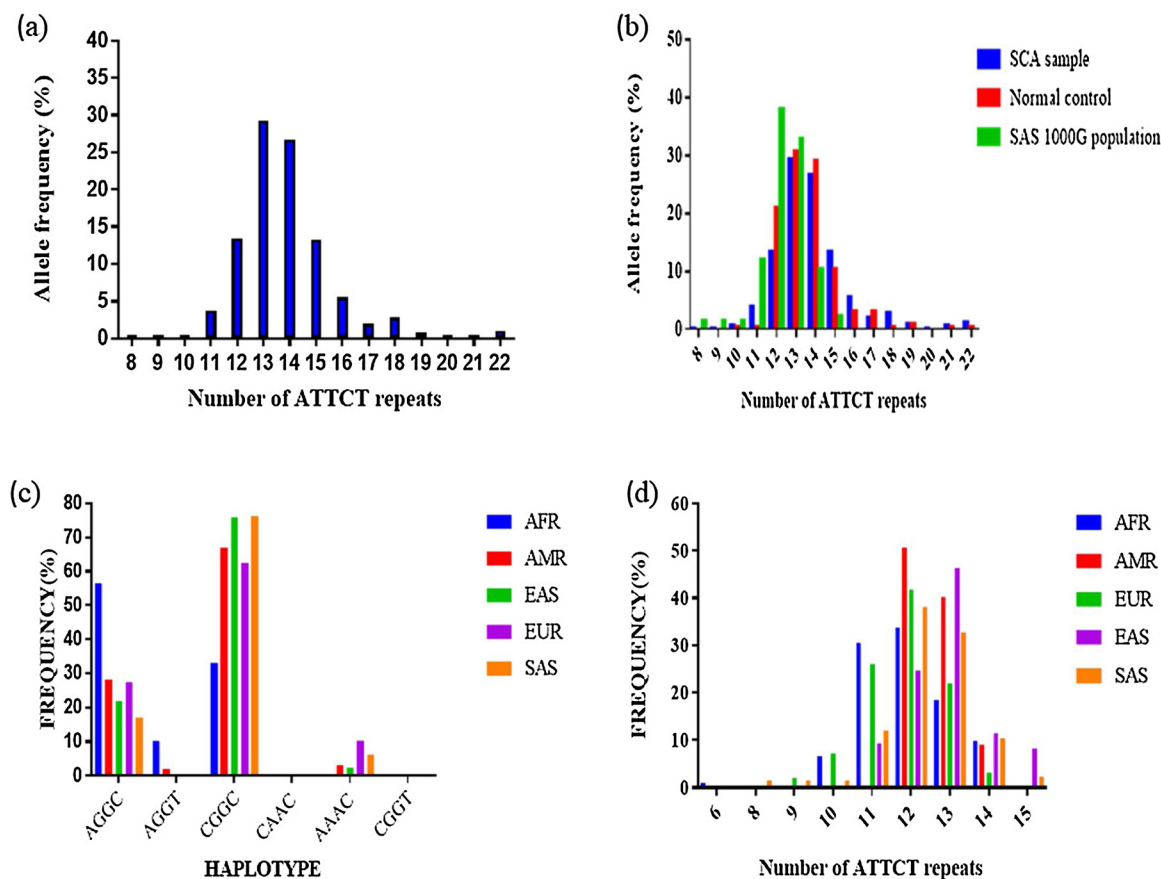
Fig. 1. Distribution of (ATTCT)n repeat and haplotype (a) Distribution of (ATTCT)n repeat in SCA patient samples (b) Comparative distribution of (ATTCT)n repeat in SCA samples, normal control and SAS reference samples (c) Reconstructed haplotype in five populations with estimated frequency (d) CGGC haplotype with repeat number in different population.

[1,6] revealed the characteristic phenotype of pure cerebellar ataxia associated with seizures. Subsequently, general features of ataxia including gait ataxia, dysarthria, variable limb ataxia, and ocular movement abnormalities have been witnessed in SCA10 positive patients [5,7,8]. Further characterization of new SCA10 families unveiled association with diverse phenotypes like polyneuropathy, pyramidal signs, cognitive impairment and neuropsychiatric disturbance, and even manifestation of ataxia without seizures [5,9]. The observed variability in the clinical spectrum of patients could be linked to the ethnicity of the reported cases [5]. In the reported cases, the age of onset of SCA10 varied from 12 to 48 years, however, progressive earlier disease onset has been anticipated in successive generations in affected members of SCA10 families [1,2,9].

Initially, SCA 10 was widely found in regions of Latin America mostly in Mexico and Brazil, where it was the second most common SCA after SCA3 (Brazil) and SCA2 (Mexico) [10]. An extensive haplotype analysis by Almeida et al., demonstrated ancestral origin of SCA10, presumed to have arisen in the Amerindian population. These findings relatively combined with the sequential dispersion of SCA10 all over Latin American countries, provide evidence for a common founder mutation [11,12]. SCA10 cases reported in patients from the USA, Mexico, Argentina, Venezuela, Colombia, Bolivia, Brazil, Peru, China and recently in Japan shared common haplotype found in the Amerindian population [5,7,13–17].

Previous reports have suggested the absence *ATXN10* pathogenic expansion in certain Asian populations like Japanese, Chinese, and Indian [14,15]. Contrary to this definition/belief, recently SCA10 positive families have been identified in the East Asian population of China and Japan. This intriguing finding prompted us to investigate the presence of SCA10 in the Indian population. Henceforth, the aim of the

present study was to investigate SCA10 (ATTCT)n repeat expansion in a reasonable cohort of 461 patients of Indian origin presenting with ADCA phenotype which were essentially negative for primary genetic screening panel for Ataxia. The distribution of *ATXN10* normal alleles was also determined in a cohort of control samples. Furthermore, the analysis and distribution of known intragenic haplotype CGGC and extended haplotype lineages flanking the pentanucleotide repeat of *ATXN10* was performed utilizing the dataset from 1000 Genome Project to support our findings through population genetics approach [18].

2. Methods

2.1. Patient recruitment

A total of 461 unrelated patients enrolled in the CSIR-GOMED project [an initiative taken for investigation of neurogenetic and other prevalent genetic disorders in the Indian population (http://gomed.igib.in)] during 2015–2018 were included in the study. These patient samples had essentially come for genetic screening for SCA and was found negative for the primary screening panel inclusive of SCA1, SCA2, SCA3, SCA6, SCA7, SCA12, and SCA17.

We also included 117 neurologically healthy subjects as controls in this study. The study was approved by the ethics committee CSIR-IGIB. Written informed consent was obtained from all patients and controls for molecular genetic testing and research.

2.2. Molecular testing for ATXN10 repeat expansion analysis

Blood samples was collected from all patients and controls included in the study. Genomic DNA was extracted from peripheral blood

Table 1
The core SCA10 haplotype and its frequency distribution with ATTCT repeats.

(ATTCT)n	6	8	9	10	11	12	13	14	15
AFR									
AGGC	3.2	1.3	1.3	8.3	30.6	35.0	16.6	3.8	0.0
AGGT	0	0	0	0	46.43	46.43	7.14	0	0
CGGC	1.1	0.0	0.0	6.5	30.4	33.7	18.5	9.8	0.0
CAAC	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
AAAC	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CGGT	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
EAS									
AGGC	0.0	0.0	0.0	0.0	14.7	17.6	29.4	14.7	23.5
AGGT	0.0	0.0	0.0	0.0	9.3	24.7	46.4	11.3	8.2
CGGC	0.0	0.0	0.0	0.0	10.1	27.0	50.6	12.4	0.0
CAAC	0	0	0	0	0	0	0	0	0
AAAC	0.0	0.0	0.0	0.0	0.0	33.3	33.3	33.3	0.0
CGGT	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AMR									
AGGC	7.1	0.0	0.0	0.0	7.1	32.1	46.4	7.1	0.0
AGGT	0.0	0.0	0.0	0.0	0.0	50.0	50.0	0.0	0.0
CGGC	0.0	0.0	0.0	0.0	0.0	50.7	40.3	9.0	0.0
CAAC	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AAAC	0.0	0.0	0.0	0.0	66.7	33.3	0.0	0.0	0.0
CGGT	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
EUR									
AGGC	0.0	0.0	0.0	9.5	23.8	28.6	31.0	4.8	2.4
AGGT	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CGGC	0.0	0.0	2.0	7.1	25.5	40.8	21.4	3.1	0.0
CAAC	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AAAC	0.0	0.0	0.0	0.0	25.0	37.5	31.3	0.0	6.3
CGGT	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
SAS									
AGGC	0.0	0.0	0.0	0.0	33.3	50.0	6.7	6.7	3.3
AGGT	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CGGC	0.0	1.5	1.5	1.5	11.9	38.1	32.8	10.4	2.2
CAAC	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AAAC	0.0	0.0	0.0	0.0	18.2	0.0	54.5	27.3	0.0
CGGT	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0

According to KGP database and Phase v2.1. EUR: European; EAS: East Asian; AMR: American; AFR: African; SAS: South Asian.

lymphocytes using modified salting out method ((Miller et al., 1988).

2.2.1. PCR and fragment analysis

Flanking PCR was performed using fluorescently labelled *ATXN*10 locus-specific primers to assess (ATTCT)n repeats in all patients and controls included in the study. The sequence of primers used were as follows: SCA10-FP (5´–6-FAM-CAGATGGCAGAATGATAAACTCAA–3´) primer and SCA10-RP (5´–AGCCTGGGCAACATAGAGAGA–3´). PCR was performed in a 10-μl reaction containing 100$_n$g of genomic DNA, 0.3 pmol of each primer, 2.5 mM of each dNTP, 15 mM MgCl$_2$, 5 M betaine and 1U of AmpliTaq Gold polymerase in the 10X buffer (Applied Biosystems). The reaction conditions were as follows: initial denaturation at 96 °C for 5 min; 35 cycles of denaturation at 96 °C for 45 s, annealing at 60 °C for 1 min, and extension at 72 °C for 3 min (all steps were completed using ABI thermal cycler). PCR products were separated on ABI 3500XL Genetic Analyzer (Applied Biosystems) and evaluated using the GeneMapper v4.1 software with ROX-500 as an internal standard marker (Applied Biosystems).

Repeat-primed (RP) PCR protocol was followed for the detection of expanded *ATXN10* allele for all samples which showed homozygous peak in the flanking PCR [19,20]. It was carried out using three primers including one FAM labelled forward primer (FP) (SCA10-5′-6-FAM-GAAGACAAATAGAAAACAGATGGCACA), a first reverse primer RP-R1 (5′- TACGCATCCCAGTTTGAGACG GAATAGAATAGAATAGAATAGAA TAGAATAGAATAGAATAG) and a second reverse RP-R2 anchor tail (5′-TACGCATCCCAGTTTGAGACG). RP-PCR was carried out as follows: 100 ng of genomic DNA, 0.8 μM FP, 0.08 μM of RP-R1 primer and 0.8 μM of the RP-R2 primer, with FAILSAFE I reagent (epicentre), and 1U of AmpliTaq Gold polymerase (Applied Biosystems). Amplified

fragments were separated and analyzed on ABI 3500XL Genetic Analyzer (Applied Biosystems) and evaluated using the GeneMapper v4.1 software (Applied Biosystems).

2.3. SCA10 haplotype analysis

SCA10 at-risk haplotype (CGGC) analysis was carried out using reported marker SNPs C(rs5764850), G(rs72556348), G(rs72556349), C(rs72556350) which are considered as stable haplotype markers in SCA10 cases flanking the repeat expansion region [12]. The study of risk haplotype variants was performed using the 1000 Genomes population data which provides information about genetic variants found in European (EUR), East Asian (EAS), American (AMR), African (AFR) and South Asian (SAS) population [18]. The genetic variants of five different populations distributed across the world were reconstructed for estimating haplotype frequency with most likely haplotype pairing among individuals of different population using PHASE v2.1 [21]. Similarly, the normal count and distribution of ATTCT repeat in five different populations with varying SNP lineages was studied using the 1000 Genomes Project and PHASE v2.1. For extended haplotype analysis in (ATTCT)n repeat flanking region, SNPs genotype variant data was obtained for *ATXN10* region covering 5118 SNPs. From these, SNPs with allele frequency > 0.05 were chosen for further haplotype analysis. A total of 56 SNPs (spanning ~26kbp region) remained for the haplotype analysis and were analyzed using PHASE v2.1.

3. Results

3.1. Brief clinical details of the study cohort

The patients included in the study was essentially diagnosed with cerebellar ataxia with variable phenotype. It included a clinically heterogeneous cohort with mild to severe apparent signs of cerebellar ataxia, dysarthria and other neurologic signs and symptoms of varying durations. Of the total samples taken for the study, 324 patients were male and 137 were female. The patients were further categorized as Early onset cerebellar ataxia (6–25 years; mean age:18.6; n = 72), adult onset cerebellar ataxia (26–45 years; mean age: 35.4 years; n = 156) and late onset cerebellar ataxia (46–80 years; mean age: 58.8 years; n = 207) with major phenotype of ADCA/SCA cases (for 6% of cases age information were not available with the referral documents).

3.2. ATTCT repeats in patients and control population

Pathogenic repeat expansion for SCA10 was not observed in any of the patients included in the study. Moreover, genetic analysis of ATTCT repeat region in *ATXN 10* gene (allele count n = 922) revealed that 78.8% of patients were heterozygous with repeats ranging from 8 to 22. The most common ATTCT repeat was 13 found in 29.27% of patients followed by 14 repeats in 26.67% of cases (Fig. 1a). Further, repeat primed PCR analysis confirmed that the remaining 21.2% of patients were homozygous with allele ranging from 11 to 19 repeats. Also, in the homozygous state, allele 13/13 ATTCT repeats were the most recurrent with the frequency of 45% followed by 14/14 repeats.

In the 117 control individuals, the ATTCT repeat ranged from 10 to 22, with 82.1% heterozygosity, and the remaining 17.9% samples were homozygous. The control group showed 13 ATTCT repeats as most frequent with an estimated frequency of 30.8% followed by 14/14 repeat. We also compared the results with the reference samples present in the 1000 Genomes Project SAS population showing 12 ATTCT repeats as the most common repeat in the cohort of 86 reference samples.

3.3. Haplotype distribution and its inference in Indian population

To assess the frequency distribution of at-risk haplotype for ATTCT expansion 'C-(ATTCT)n-GGC', the genotype variants of SNPs

Table 2
Extended Haplotype flanking CGGC-SCA10 core haplotype structure. The text in italics represents haplotype lineage specifically observed in AMR or EAS population. The other haplotype in Boxed manner represents prevalent haplotype 1-CGGC-A and 2-CGGC-B.

AFR	AMR	EAS	EUR	SAS	ATTCTn Allele	ATTCTn length	Haplo-5'	5'ARM	SCA10 core haplotype	Haplo-3'	3'arm
0	*1*	*0*	*0*	*0*	*8*	*14*	*TTGGGCCTTGTGTGTCAGAAAAAAAAAGGG*	*16*	*CGGC*	*TCACGTGTGAAGAGTAGCCCAACC*	*A*
0	*1*	*0*	*0*	*0*	*7*	*13*	*TTGAACTATGCGTCTCAAAAAAAAGAG*	*13*	*CGGC*	*TCACGTGTGAAGATTAGCCGAACC*	*B*
0	*0*	*1*	*0*	*0*	*8*	*14*	*TTGGACCTTGTGTCAGAAAAAAAAAGAG*	*4*	*CGGC*	*TCACGTGTGAAGAGTAGCCGAACC*	*C*
0	*0*	*2*	*0*	*0*	*7*	*13*	*TTGGGCCTCGTGTCAGAAAGAAAGAG*	*8*	*CGGC*	*TCACATCTGAAGATTGGCCGAACC*	*D*
0	*0*	*1*	*0*	*0*	*7*	*13*	*TTGGGCCATGCGTCAGAAAGGGAGGGG*	*18*	*CGGC*	*TCACGTGCGCGGAGTAGCCCAACC*	*H*
0	*0*	*1*	*0*	*0*	*7*	*13*	*TTGGGCCATGCGTCAGAAAGGGAGGGG*	*18*	*CGGC*	*TCACGTGTGCGGAGTAGCCCAACC*	*L*
1	0	0	0	0	1	6	TTGGGCCTTGTGTCAGAAAAAAAAGAG	1	CGGC	TCACGTGTGAAGAGTAGCCCAACC	A
0	0	0	0	1	2	8	TTGGGCCTTGTGTCAGAAAAAAAAGAG	1	CGGC	TCACGTGTGAAGAGTAGCCCAACC	A
0	0	0	2	1	3	9	TTGGGCCTTGTGTCAGAAAAAAAAGAG	1	CGGC	TCACGTGTGAAGAGTAGCCCAACC	A
4	0	0	6	1	4	10	TTGGGCCTTGTGTCAGAAAAAAAAGAG	1	CGGC	TCACGTGTGAAGAGTAGCCCAACC	A
24	10	1	17	10	5	11	TTGGGCCTTGTGTCAGAAAAAAAAGAG	1	CGGC	TCACGTGTGAAGAGTAGCCCAACC	A
28	24	7	23	33	6	12	TTGGGCCTTGTGTCAGAAAAAAAAGAG	1	CGGC	TCACGTGTGAAGAGTAGCCCAACC	A
11	20	15	6	16	7	13	TTGGGCCTTGTGTCAGAAAAAAAAGAG	1	CGGC	TCACGTGTGAAGAGTAGCCCAACC	A
3	3	3	2	7	8	14	TTGGGCCTTGTGTCAGAAAAAAAAGAG	1	CGGC	TCACGTGTGAAGAGTAGCCCAACC	A
0	0	2	0	1	9	15	TTGGGCCTTGTGTCAGAAAAAAAAGAG	1	CGGC	TCACGTGTGAAGAGTAGCCCAACC	A
2	0	0	0	0	5	11	TTGGGCCATGTGTCAGAAAGAAAGAG	5	CGGC	TCACGTGTGAAGAGTAGCCCAACC	A
0	1	0	1	0	5	11	TTGGGCCATGTGTCAGAAAAAAAAGAG	6	CGGC	TCACGTGTGAAGAGTAGCCCAACC	A
1	0	0	1	0	6	12	TTGGGCCATGTGTCAGAAAAAAAAGAG	6	CGGC	TCACGTGTGAAGAGTAGCCCAACC	A
1	0	0	0	0	7	13	TTGGGCCATGTGTCAGAAAAAAAAGAG	6	CGGC	TCACGTGTGAAGAGTAGCCCAACC	A
5	0	0	0	0	8	14	TTGGGCCATGTGTCAGAAAAAAAAGAG	6	CGGC	TCACGTGTGAAGAGTAGCCCAACC	A
2	0	0	0	0	6	12	TTGGGCCTTGTGTCAGAAAGAAAGAG	11	CGGC	TCACGTGTGAAGAGTAGCCCAACC	A
1	0	0	0	0	8	14	TTGGGCCTTGTGTCAGAAAGAAAGAG	11	CGGC	TCACGTGTGAAGAGTAGCCCAACC	A
0	0	1	0	0	6	12	TTGGGCCTCGTGTCAGAAAAAAAAGAG	14	CGGC	TCACGTGTGAAGAGTAGCCCAACC	A
0	0	0	1	0	6	12	TTGGGCCTTGTGTCAGAAAAAAAAGGG	16	CGGC	TCACGTGTGAAGAGTAGCCCAACC	A
0	1	0	2	0	7	13	TTGGGCCTTGTGTCAGAAAAAAAAGGG	16	CGGC	TCACGTGTGAAGAGTAGCCCAACC	A
0	0	0	0	2	8	14	TTGGGCCTTATGTCAGAAAAAAAAGAG	20	CGGC	TCACGTGTGAAGAGTAGCCCAACC	A
1	0	0	0	0	7	13	TTGGGCCTTGTGCCAGAAAAAAAAGAG	22	CGGC	TCACGTGTGAAGAGTAGCCCAACC	A
0	0	0	0	1	2	8	TTGGACTTTATGTCAGAAAAAAAAGAG	2	CGGC	TCACGTGTGAAGATTAGCCGAACC	B
0	0	0	0	1	3	9	TTGGACTTTATGTCAGAAAAAAAAGAG	2	CGGC	TCACGTGTGAAGATTAGCCGAACC	B
1	0	0	1	0	4	10	TTGGACTTTATGTCAGAAAAAAAAGAG	2	CGGC	TCACGTGTGAAGATTAGCCGAACC	B
1	6	7	5	4	5	11	TTGGACTTTATGTCAGAAAAAAAAGAG	2	CGGC	TCACGTGTGAAGATTAGCCGAACC	B
0	10	15	12	11	6	12	TTGGACTTTATGTCAGAAAAAAAAGAG	2	CGGC	TCACGTGTGAAGATTAGCCGAACC	B
2	6	24	12	23	7	13	TTGGACTTTATGTCAGAAAAAAAAGAG	2	CGGC	TCACGTGTGAAGATTAGCCGAACC	B
0	2	7	1	5	8	14	TTGGACTTTATGTCAGAAAAAAAAGAG	2	CGGC	TCACGTGTGAAGATTAGCCGAACC	B
0	0	6	0	2	9	15	TTGGACTTTATGTCAGAAAAAAAAGAG	2	CGGC	TCACGTGTGAAGATTAGCCGAACC	B
1	0	0	0	0	4	10	TTGGGCTATGTGTCAGAAAAAAAAGAG	3	CGGC	TCACGTGTGAAGATTAGCCGAACC	B
0	1	0	0	0	5	11	TTGGGCTTTGTGTCAGAAAAAAAAGAG	7	CGGC	TCACGTGTGAAGATTAGCCGAACC	B
1	0	0	0	0	6	12	TTGGGCTTTGTGTCAGAAAAAAAAGAG	7	CGGC	TCACGTGTGAAGATTAGCCGAACC	B
0	0	0	2	0	5	11	TTGGGCTTTATGTCAGAAAAAAAAGAG	9	CGGC	TCACGTGTGAAGATTAGCCGAACC	B
0	0	0	1	0	6	12	TTGGGCTTTATGTCAGAAAAAAAAGAG	9	CGGC	TCACGTGTGAAGATTAGCCGAACC	B
0	1	0	1	0	6	12	TTGAATATGCGTCTCAAAAAAAAGAG	13	CGGC	TCACGTGTGAAGATTAGCCGAACC	B
0	0	1	0	0	6	12	TTGGACTATATGTCAGAAAAAAAAGAG	15	CGGC	TCACGTGTGAAGATTAGCCGAACC	B
0	0	1	1	0	7	13	TTGGACTATATGTCAGAAAAAAAAGAG	15	CGGC	TCACGTGTGAAGATTAGCCGAACC	B
0	0	0	0	1	6	12	TTGGGCCTTATGTCAGAAAAAAAAGAG	20	CGGC	TCACGTGTGAAGATTAGCCGAACC	B
0	0	0	0	1	6	12	TTGGACTTTATGTCAGAAAAAAAAGGG	21	CGGC	TCACGTGTGAAGATTAGCCGAACC	B
0	0	0	0	1	6	12	TTGGGCCTTGTGTCAGAAAAAAAAGAG	1	CGGC	TCACGTGTGAAGAGTAGCCGAACC	C
0	0	1	0	0	7	13	TTGGACTTTATGTCAGAAAAAAAAGAG	2	CGGC	TCACGTGTGAAGAGTAGCCGAACC	C
0	0	0	0	1	4	10	TTGGACCTTGTGTCAGAAAAAAAAGAG	4	CGGC	TCACGTGTGAAGAGTAGCCGAACC	C
0	0	1	0	0	5	11	TTGGGCCTCGTGTCAGAAAGAAAGAG	8	CGGC	TCACATCTGAAGATTGGCCGAACC	D
0	0	0	0	1	5	11	CCGAGCCACGTGTCAGAAAAAAAAGAG	10	CGGC	TCACATCTGAAGAGTGTTGCGGTT	E
1	0	0	0	0	6	12	TTGGACTTTACGTTAGAAAAGGAGGAG	12	CGGC	TCACGTGTGAAGAGTAGCCCAACC	F
0	0	0	1	0	6	12	CCGAGCCACGTGTCACAGGGAAAAGAG	17	CGGC	TCACGTGTGAAGAGTGGTGCGGTT	G
0	0	0	0	1	6	12	TTGGGCCATGCGTCAGAAAGGGAGGGG	18	CGGC	TCACGTGCGCGGAGTAGCCCAACC	H
0	0	0	0	1	6	12	TTGGGCCTTGCATCTGAAAAAAAAGAG	19	CGGC	GCACGTGTGAAGAGTAGCCCAACC	I
0	0	0	0	1	7	13	TTGGGCCTTGCATCTGAAAAAAAAGAG	19	CGGC	GCACGTGTGAAGAGTAGCCCAACC	I
0	0	0	0	1	7	13	TTGGACCATATGTCAGAAAAAAAAGAG	24	CGGC	GCACGTGTGAAGAGTAGCCCAACC	I
0	0	0	2	0	6	12	TTGGGCCTTGTGTCAGAAAAAAAAGAG	1	CGGC	TCACGTCTGAAGAGTAGCCCAACC	J
1	0	0	0	0	7	13	TTAGGGCTTGCGTCAGGAAAAAGAAAG	23	CGGC	TCAGGTGTGAAGGGCAGCCCAACC	K
0	0	0	0	1	7	13	TTGGGCCTTGTGTCAGAAAAAAAAGAG	1	CGGC	TCACATCTGAAGATTAGCCGAACC	M
0	0	0	0	1	7	13	TTGGACTTCATGTCAGAAAGAAAGAG	25	CGGC	TCACGTGTGAAGATTGGCCGAACC	N

rs5764850, rs72556348, rs72556349, and rs72556350 along with the ATTCT repeats from 1000 Genomes Population database were used. Haplotype construction with most likely haplotype pairing in European, East Asian, American, African and South Asian population using PHASE v2.1 revealed six possible haplotype pairing. The possible pair with common SNPs found was CGGC, AGGC, AGGT, CAAC, AAAC, and CGGT (Table 1). The CGGC (spanning over 2 kbp) haplotype was most common among all the populations followed by AGGC whereas another haplotype variant was found in specific population only as shown in

(Fig. 1c). Amongst five different populations, estimated frequency of CGGC haplotype was found to be 33% in AFR, 67% in AMR, 75.7% in EAS, 62.3% in EUR and 76.13% in SAS population. The CGGC haplotype was not found to be segregating with either large normal or small normal ATTCT repeat length. The distribution of ATTCT repeat number with CGGC haplotype in five populations demonstrated that 12 repeats of ATTCT was most frequently associated with CGGC haplotype across all populations including the SAS population (Fig. 1d). As the at-risk haplotype lineage was found to be quite frequent, we investigated the

Table 3
The comparative studies of SCA10 in different world populations.

Population	Number of patients	Age of onset (years)	Number of ATTCT repeat Controls	Number of ATTCT repeats In Patients	References
Brazil	90	35.5 (22-46)	10 - 29	850-1500	[13,24]
Mexico	19	26.7 (14-44)	10 - 22	920 - 4,140	[7]
Argentina	5	35	10 - 29	1100	[22]
Peru	1	47	9 - 32	1,760	[16]
Minnesota	1	89	< 32	1400	[29]
Bolivia	2	38-60	?	1611	[23]
China	1	45	9 - 32	852	[15]
Japan	2	34-65		1500	[14]
Cypriot	58	20-50	11 - 20	10–19	[26]
German	440	19-58	10 - 22	15-16	[25]
White American, French-Canadian, Italian, Japanese & Spanish	478	?	10 - 22	10-21	[27]
French	123 families	?	11 - 22	11 - 22	[11]
India	461	43 (6-80)	10-22	8-22	Present study

differences in at-risk haplotype sub lineage, spanning the repeat region of the different populations. We studied the genetic variants in an extended region surrounding ATTCT repeats. The extended haplotype analysis showed that the neighbouring haplotype arm surrounding the at-risk haplotype CGGC' majorly comprised of haplotype-A, 1-CGGC-A (55%) (Table 2) and haplotype-B, 2-CGGC-B (32%) and were present across all five sub-population of KGP dataset. As SCA10 is known to occur only in people of Amerindian/Latin America and East Asian Ancestry (China and Japan), we looked across any diverged haplotype lineage with ATTCT > 13 repeats specific to AMR and EAS population. On further analysing the 5′ neighbouring divergent arm, we observed haplotype lineages (comprising of different SNPs) in either AMR or EAS populations. Furthermore, none of the haplotype lineages had shown presence in both the populations together, suggesting that these haplotype lineages are population specific as well as they were absent in other KGP (EUR, SAS and AFR) populations.

4. Discussion

SCA10 is one of the rare ADCA caused by pentanucleotide repeat expansion in the intron 9 of the *ATXN10* gene. Till date, SCA10 cases have been reported in several countries including, USA, Mexico, Argentina, Colombia, Bolivia, Brazil and Peru (Table 3) [5,7,13,16,17,22–24]. The reported cases share common SNPs associated with the SCA10 expansion locus proposing that SCA10 in Latin America has a common origin coming from an Amerindian populace. Recently, two SCA10 affected families has been documented in China and Japan having ATTCT expansion sharing common haplotype within the repeat region [14,15].

Genetic testing of SCA10 in a substantial cohort of Indian SCA patients did not show any pathological repeat expansion in the *ATXN10* gene. Similar studies on SCA10 have documented the absence of pathological repeat expansion in the Cyprus and German populations [25,26]. In the current study, the ATTCT repeat length was found to fall between 8 to 22 repeats which is the non-pathological or normal range. It may be suggested that the near absence of SCA10 in Indian population could possibly be due to differing ethnicities and might manifest in populations with specific genetic backgrounds [27]. Additionally to strengthen our hypothesis of absence of SCA10 in the Indian population, haplotype analysis flanking the ATTCT expansion region was carried out across various ethnic population in 1000 Genomes data (KGP). The study revealed that the at-risk haplotype is comparatively more frequent in all populations with large normal or small normal ATTCT length. However, extended haplotype analysis showed some diversified lineage at 5′ end of the CGGC haplotype with ATTCT > 13 repeats. This lineage was specifically observed in the American or East Asian population excluding other populations in KGP dataset. As

common risk haplotype has been observed across all populations, so we can't rule down the possible hypothesis that independent expansion events on typically shared haplotype lineage but in different geographical regions may have possibly happened by chance event or by the backflow of human migrants from Beringia to other parts of the world [15,28]. Altogether, these facts point towards the absence of SCA10 in Indian population and haplotype-dependent assessment furthermore suggests its occurrence to be rare in South Asian, European and African population.

Although the rarity of SCA10 has already been demonstrated previously, we attempted to investigate a large cohort of Ataxia patients for SCA10 keeping in view the recently identified cases from China and Japan. Accordingly, this study does not target SCA patients manifesting the core features of SCA10. However, we adopted a large sample size to reduce the false negatives in view of phenotypic heterogeneity to overcome the limitation. We acknowledge that patients specifically exhibiting SCA10 features may give the correct scenario of its presence in the Indian population. Another major limitation may be stated as the study being hospital based rather than a population level survey to document the exact absence of SCA10 from India. However, our being the largest single centre in this context, catering to major tertiary care centres which are accessible to patients far and wide. Yet, SCA10 cases with precise clinical phenotype may be missed or go undiagnosed keeping in view the diverse ethnicity and geographical background prevailing in India.

5. Conclusion

This study establishes the rarity of SCA10 in the unrelated Indian spinocerebellar ataxia patients with heterogeneous ethnicity. It is further confirmed by population genetics based inferences suggesting it be rare or absent and might manifest in populations with specific genetic background from where it originated. However, this study provides future references while considering the diagnostic work-up of SCAs in the Indian population and to decide upon inclusion of SCA10 in panel based genetic investigations even for Indians living abroad. This approach may also be adopted for the screening of other subtypes of SCAs, i.e. other rare SCAs e.g. SCA31, SCA36, and SCA37.

Tech:MLP1802 projects.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgment

References

[1] L. Zu, K.P. Figueroa, R. Grewal, S. Pulst, Mapping of a new autosomal dominant spinocerebellar Ataxia to chromosome 22, Am. J. Hum. Genet. 64 (1999) 594–599, https://doi.org/10.1086/302247.

[2] T. Matsuura, T. Yamagata, D.L. Burgess, A. Rasmussen, R.P. Grewal, K. Watase, M. Khajavi, A.E. McCall, C.F. Davis, L. Zu, M. Achari, S.M. Pulst, E. Alonso, J.L. Noebels, D.L. Nelson, H.Y. Zoghbi, T. Ashizawa, Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10, Nat. Genet. 26 (2000) 191–194, https://doi.org/10.1038/79911.

[3] E. Pennisi, ENCODE project writes eulogy for junk DNA, Science 337 (2012) 11591161, , https://doi.org/10.1126/science.337.6099.1159.

[4] I. Alonso, L.B. Jardim, O. Artigalas, M.L. Saraiva-Pereira, T. Matsuura, T. Ashizawa, J. Sequeiros, I. Silveira, Reduced penetrance of intermediate size alleles in spino-cerebellar spinocerebellar ataxia type 10, Neurology 66 (2006) 1602–1604, https://doi.org/10.1212/01.wnl.0000216266.30177.bb.

[5] H.A.G. Teive, R.P. Munhoz, W.O. Arruda, S. Raskin, L.C. Werneck, T. Ashizawa, Spinocerebellar ataxia type 10 - A review, Park. Relat. Disord. 17 (2011) 655–661, https://doi.org/10.1016/j.parkreldis.2011.04.001.

[6] T. Matsuura, M. Achari, M. Khajavi, L.L. Bachinski, H.Y. Zoghbi, T. Ashizawa, Mapping of the gene for a novel spinocerebellar ataxia with pure cerebellar signs and epilepsy, Ann. Neurol. 45 (1999) 407–411, https://doi.org/10.1002/1531-8249(199903)45:3<407 AID-ANA21 > 3.0.CO;2-D..

[7] A. Rasmussen, T. Matsuura, L. Ruano, P. Yescas, A. Ochoa, T. Ashizawa, E. Alonso, Clinical and genetic analysis of four Mexican families with spinocerebellar ataxia type 10, Ann. Neurol. 50 (2001) 234–239, https://doi.org/10.1002/ana.1081.

[8] H.A.G. Teive, T. Ashizawa, Spinocerebellar ataxia type 10: from amerindians to latin americans, Curr. Neurol. Neurosci. Rep. 13 (2013) 9–11, https://doi.org/10.1007/s11910-013-0393-9.

[9] X. Lin, T. Ashizawa, Recent progress in spinocerebellar ataxia type-10 (SCA10), Cerebellum 4 (2005) 37–42, https://doi.org/10.1080/14734220510007897.

[10] H.A.G. Teive, B.B. Roa, S. Raskin, P. Fang, W.O. Arruda, Y.C. Neto, R. Gao, L.C. Werneck, T. Ashizawa, Clinical phenotype of Brazilian families with spino-cerebellar ataxia 10, Neurology 63 (2004) 1509–1512, https://doi.org/10.1212/01.WNL.0000142109.62056.57.

[11] A.C. Jackson, M. Melanson, J.P. Rossiter, Familial Herpes Simplex Encephalitis, (2002), pp. 406–409, https://doi.org/10.1002/10123.

[12] T. Almeida, I. Alonso, S. Martins, E.M. Ramos, L. Azevedo, K. Ohno, A. Amorim, M.L. Saraiva-Pereira, L.B. Jardim, T. Matsuura, J. Sequeiros, I. Silveira, Ancestral origin of the ATTCT repeat expansion in spinocerebellar ataxia type 10 (SCA10), PLoS One 4 (2009) 1–6, https://doi.org/10.1371/journal.pone.0004553.

[13] H. Afonso, G. Teive, A. Moro, M. Moscovich, W.O. Arruda, R. Puppi, S. Raskin, G. Mary, G. Teive, N. Dallabrida, T. Ashizawa, Spinocerebellar Ataxia Type 10 in the South of Brazil : The Amerindian-Belgian Connection, (2015), https://doi.org/10.1590/0004-282X20150086 2014–2016.

[14] H. Naito, T. Takahashi, M. Kamada, H. Morino, H. Yoshino, N. Hattori, H. Maruyama, H. Kawakami, M. Matsumoto, First report of a Japanese family with spinocerebellar ataxia type 10: the second report from Asia after a report from China, PLoS One 12 (2017) 1–11, https://doi.org/10.1371/journal.pone.0177955.

[15] K. Wang, K.N. Mcfarland, J. Liu, D. Zeng, I. Landrian, G. Xia, Y. Hao, M. Jin, C.J. Mulligan, W. Gu, T. Ashizawa, Spinocerebellar Ataxia type 10 in chinese han, Neurol. Genet. 1 (2015) 1–3 doi:10.1212/ NXG.0000000000000026.

[16] L. Leonardi, C. Marcotulli, K.N. McFarland, A. Tessa, R. DiFabio, F.M. Santorelli, F. Pierelli, T. Ashizawa, C. Casali, Spinocerebellar ataxia type 10 in Peru: the missing link in the Amerindian origin of the disease, J. Neurol. 261 (September (9)) (2014) 1691–1694, https://doi.org/10.1007/s00415-014-7394-8.

[17] K. Bushara, K.N. McFarland, I. Landrian, D. Hutter, H.A.G. Teive, A. Rasmussen, C.J. Mulligan, T. Ashizawa, Expansion of the spinocerebellar ataxia type 10 (SCA10) repeat in a patient with Sioux Native American ancestry, PLoS One 8 (2013) 8–11, https://doi.org/10.1371/journal.pone.0081342.

[18] 1000 Genomes Project Consortium, A global reference for human genetic variation, Nature 526 (2015) 68, https://doi.org/10.1038/nature15393.

[19] C. Cagnoli, C. Michielotto, T. Matsuura, T. Ashizawa, R.L. Margolis, S.E. Holmes, C. Gellera, N. Migone, A. Brusco, Detection of large pathogenic expansions in FRDA1, SCA10, and SCA12 genes using a simple fluorescent repeat-primed PCR assay, J. Mol. Diagn. 6 (2004) 96–100, https://doi.org/10.1016/S1525-1578(10)60496-5.

[20] J.P. Warner, L.H. Barron, D. Goudie, K. Kelly, D. Dow, D.R. Fitzpatrick, D.J. Brock, A general method for the detection of large CAG repeat expansions by fluorescent PCR, J. Med. Genet. 33 (1996) 1022–1026, https://doi.org/10.1136/JMG.33.12.1022.

[21] M. Stephens, N.J. Smith, P. Donnelly, A new statistical method for haplotype re-construction from population data, Am. J. Hum. Genet. 68 (2001) 978–989, https://doi.org/10.1086/319501.

[22] E.M. Gatto, R. Gao, M.C. White, M.C. Uribe Roca, J.L. Etcheverry, G. Persi, J.J. Poderoso, T. Ashizawa, Ethnic origin and extrapyramidal signs in an argenti-nean spinocerebellar Ataxia type 10 family, Neurology 69 (2007) 216–218, https://doi.org/10.1212/01.wnl.0000265596.72492.89.

[23] J.F. Baizabal-Carvallo, G. Xia, P. Botros, J. Laguna, T. Ashizawa, J. Jankovic, Bolivian kindred with combined spinocerebellar ataxia types 2 and 10, Acta Neurol. Scand. 132 (2015) 139–142, https://doi.org/10.1111/ane.12371.

[24] P. De Carvalho Aguiar, T. Fuchs, V. Borges, K.M. Lamar, S.M.A. Silva, H.B. Ferraz, L. Ozelius, Screening of Brazilian families with primary dystonia reveals a novel THAP1 mutation and a de novo TOR1A GAG deletion, Mov. Disord. 25 (2010) 2854–2857, https://doi.org/10.1002/mds.23133.

[25] G. Aydin, G. Dekomien, S. Hoffjan, J.T. Epplen, L. Arning, Frequency of SCA8, SCA10, SCA12, SCA36, FXTAS and C9orf72 repeat expansions in SCA patients negative for the most common SCA subtypes, BMC Neurol. 18 (2018) 1–8, https://doi.org/10.1186/s12883-017-1009-9.

[26] C. Votsi, E. Zamba-papanicolaou, A. Georghiou, T. Kyriakides, S. Papacostas, K.A. Kleopa, M. Pantzaris, K. Christodoulou, Journal of the Neurological Sciences Investigation of SCA10 in the Cypriot population : further exclusion of SCA dynamic repeat mutations, J. Neurol. Sci. 323 (2012) 154–157, https://doi.org/10.1016/j.jns.2012.09.006.

[27] T. Matsuura, L.P.W. Ranum, V. Volpini, Spinocerebellar Ataxia Type 10 Is Rare in Populations Other Than Mexicans the Online Version of This Article, Along With Updated Information and Services, Is Located on the World Wide Web at : Clinical / Scientific Notes Spinocerebellar Ataxia Type, 10 i (2012), https://doi.org/10.1212/WNL.58.6.983.

[28] C.J. Mulligan, A. Kitchen, M.M. Miyamoto, Updated three-stage model for the peopling of the Americas, PLoS One 3 (2008) e3199doi:0.1371/journal.-pone.0003199.

[29] J. Liu, K.N. McFarland, I. Landrian, S.S. Wu, M. Bower, D. Hutter, K. Bushara, H.A.G. Teive, T. Ashizawa, Identifying novel interruption motifs in spinocerebellar ataxia type 10 expansions, Neurol. Clin. Neurosci. 2 (2014) 38–43, https://doi.org/10.1111/ncn3.78.