Integrated whole exome sequencing and functional approach delineate genetic heterogeneity in cerebellar ataxias

Renu Kumari
CSIR-Institute of Genomics and Integrative Biology (CSIR -IGIB)
Bharath Ram Uppilli
CSIR-Institute of Genomics and Integrative Biology (CSIR -IGIB)
Sunil Shakya
All India Institute of Medical Sciences
Ajay Garg
All India Institute of Medical Sciences
Aditi Joshi
CSIR-Institute of Genomics and Integrative Biology (CSIR -IGIB)
Afreen Khan
CSIR-Institute of Genomics and Integrative Biology (CSIR -IGIB)
Varun Suroliya
All India Institute of Medical Sciences
Akhilesh K Sonakar
All India Institute of Medical Sciences
Pooja Sharma
CSIR-Institute of Genomics and Integrative Biology (CSIR -IGIB)
Inder Singh
All India Institute of Medical Sciences
Divya KP
Sree Chitra Tirunal Institute of Medical Sciences and Technology
Mitali Mukerji
CSIR-Institute of Genomics and Integrative Biology (CSIR -IGIB)
Achal K Srivatava
All India Institute of Medical Sciences
Mohammed Faruq (faruq.mohd@igib.in)
CSIR-Institute of Genomics and Integrative Biology (CSIR -IGIB)

Research Article

Keywords: Cerebellar Ataxias, Whole exome sequencing, SPTB, ataxia genetics, India

Posted Date: August 2nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-757408/v1

Abstract

Purpose

Disease deconvolution in heterogeneous cerebellar ataxias (CAs) needs a focussed approach to overcome the diagnostic challenges. A diverse clinical presentation with over 100 reported genetic loci, in addition to the various challenges associated with genotype-phenotype correlation complicate the genetic diagnosis in 40-60% of the CA cases that remain uncharacterized. We present here an integrated whole exome sequencing combined with a functional validation approach to delineate the genetic etiology in Indian CA patients.

Method

A total of 50 familial and sporadic progressive CA families (negative for CNG expansion) including 101 subjects were recruited for this study. Index patients from 50 families were subjected to singleton whole exome sequencing (S-WES). Family-based WES (F-WES) was carried out for seven S-WES selected families. Protein simulation and docking studies were performed for seven genetic variants identified through WES. A Cell line-based model was used to assess disease signatures for variants in *KCNC3* and a new candidate gene, *SPTB*.

Results

Clinically relevant variants identified in 70% (35/50) of the selected families. We achieved a 50% (25/50) definitive diagnostic yield and 14% (7/50) probable diagnostic yield while 6% (3/50) of the families showed variants of uncertain significance. We prioritized compound heterozygous variants in a candidate gene, *SPTB* for cerebellar ataxia with hereditary spherocytosis. Lymphoblastoid cell line derived from a patient with a *KCNC3* variant showed altered disease signatures with induced ROS and elevated unfolded protein response markers at the basal level.

Conclusion

Our results highlight an extensive experimental design for the genetic diagnosis of CA. Through targeted analysis of ataxia phenotype-derived gene panel in S-WES, new gene identification through F-WES, and revaluation of unsolved families' WES data, we identified novel, reported and other clinically relevant variants in CA patients. Bioinformatic protein modeling along with the cellular insights into the pathogenicity of novel variants enabled delineation of genetic diagnostics and enhanced the mechanistic understanding of CAs.

Introduction

Cerebellar ataxias (CAs) are a group of rare monogenic neurodegenerative disorders with heterogeneous clinical presentation and variable age at onset. The debilitating disease progression and disability

imposed by CAs on the patients impose an economic burden to the affected families along with a disease burden for future generations.

Amongst the clinical manifestations of CAs, incoordination of voluntary motor movements is a cardinal cerebellar symptom besides more than 35 other non-cerebellar and other non-neuronal symptoms [1]. While more than 100 different genetic loci have been implicated in CA subtypes so far, ataxia is also present as a clinical feature in over 500 other neurological disorders with known genetic etiology [2–4]. Heterogeneous clinical presentation, a large number of underlying variants, anticipation, pleiotropy, variable expressivity, and a highly diverse population structure are some of the factors that obscure a differential or definite diagnosis for CA subtypes.

For any monogenic disorder, a rapid genetic diagnosis is an invaluable aid for the healthcare system as it enables timely disease prognosis and better patient management. Among the genetically defined CA cases, tandem expansions of nucleotide repeats account for nearly 30-40% of the disease burden, while conventional variants add another half to the causal factor list [5]. Whole exome sequencing (WES) as a high throughput tool for detection of all coding variants, has been found to be more efficient than other genetic tools at improving diagnostic yield and enabling novel gene discoveries [6–8]. Our previous study in autosomal recessive CA (ARCA) families has enabled us to identify both, already reported as well as novel variants in typical and atypical ataxia genes with a much higher diagnostic yield (56%) using WES, as compared to targeted resequencing [9,10]. WES could therefore be a more appropriate tool for genetic delineation in otherwise unresolved cases of CA families, specifically the sporadic CA occurrences.

The present study applied a rapid, focussed, and integrated approach involving genetic and functional analysis for the diagnosis of CAs. Use of WES in indexed patients (Singleton; S-WES) and family members (family-based design; F-WES) in 50 families enabled us to delineate clinically relevant genetic defects in 70% (35/50) of the selected families with a definitive genetic diagnosis in 50% (25/50) of the families.

We observed genetic variants in CAs, hereditary 1spastic paraplegia (HSP), and other neurological disorders genes. Moreover, we propose a new candidate gene *SPTB* for CAs with hereditary spherocytosis. Through our functional approach of testing pathogenicity of identified variants with rapidity and sensitivity, we demonstrated the potential of patient-derived lymphoblastoid cell lines (LCLs) based assays for elucidation of neurodegeneration pathways, utilizing *KCNC3* variant as a model.

Results

Clinical representation of selected cases

Amongst the 54 uncharacterized cases selected from the 50 families, the gender ratio (male (M): female (F)) was 38:16. The mean (SD) age of onset (AO) of the CA subtypes in the families was observed to be 35.6 (11.6) years for autosomal dominant cerebellar ataxias (ADCA), 18.2 (9.4) years for sporadic early-

onset cerebellar ataxias (SPEOCA), and 51.2(6.3) years in case of sporadic late-onset cerebellar ataxias (SPLOCA) (Supplementary Table S1). Among the cerebellar features of CA, gait was the most prevalent cerebellar manifestation in patients exhibited by 94% of the cases, followed by dysarthria (85%), intention tremor (70%), and nystagmus (56%). A pyramidal sign such as the extensor plantar reflexes (areflexia, hyperreflexia, and clonus), tone (spasticity), and planter (Babinski sign) was presented by 30-40% of the families. Extrapyramidal signs (postural tremor, rigidity) as a subclinical feature were represented in very few of the evaluated patients. Other neurological signs such as peripheral neuropathy, autonomic dysfunction, skeletal abnormality, and psychiatric symptoms were also presented by the cases (Supplementary Fig. S1 and Supplementary Table S2).

Diagnostic yield

We used S-WES in 50 familial and sporadic families of CA to screen variants in gene panels consisting of ataxia phenotype genes. Further, to enhance the diagnostic yield and enabling new gene discoveries, we applied F-WES in seven out of the 50 families. Technical validation, segregation analysis (Supplementary Table S3 and S4), frequency estimation (in 257 ethnically matched controls and in-house data), protein modelling, and genotype-phenotype correlation were performed for the prioritized variants (Figure 1).

We identified and designated clinically relevant variants in 70% of the families (35/50), 62% with S-WES and 8% through F-WES. Among the categorized families, defined genetic etiology in ADCA, SPEOCA, and SPLOCA was 57%, 86%, and 57% respectively. We delineated definitive genetic diagnosis in 50% of the families (25/50). Reanalysis using updated annotation pipelines and data resources enabled further identification of probable diagnostic variants in 14 % (7/50) of the families (Table 1).

While following the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG-AMP) guidelines for variant classification [11–13] in both, the S-WES and the F-WES design, we identified 41 clinically relevant variants in 29 genes including 12 pathogenic, 25 likely pathogenic and 4 variants of uncertain significance (VUS). The above-mentioned 41 variants also include 10 novels, 12 reported (Clinvar and/or HGMD), and 19 rare variants. (Table 2 and Supplementary file 1). Of these, 37 pathogenic and likely pathogenic variants were seen in 32 families and 4 VUS variants in three families (Table 2).

Spectrum of genetic variants in cerebellar ataxia subtypes

Following WES in single and family-based design, we discovered 14 CA subtypes in our cohort, including 11 spinocerebellar ataxia (SCA) and 3 spinocerebellar ataxia autosomal recessive (SCAR) subtypes in 17 families (Table 2 and Supplementary file 1).

Among the SCA subtype variants observed, we found three previously reported pathogenic variants and two likely pathogenic novel variants. In AT1798, we observed a novel heterozygous (Het) variant, p.Glu53Asp in *EEF2* (SCA26, 609306), which lies in proximity to the phosphorylation site, p.Thr56. Two phosphorylation sites, p.Thr56 and p.Ser595 regulate the eEF2 role in the cell cycle and translational

machinery. Variant p.Ser595 directly regulates p.Thr56 and is required for its efficient phosphorylation. Our reported variant p.Pro596His, which lies next to the phosphorylation site p.Ser595, leads to impaired translation and increases the susceptibility for proteostatic disruption. Identified variant p.Glu53Asp lies in a GTP binding domain (P-loop containing nucleoside triphosphate hydrolase) next to phosphorylation site, p.Thr56, and probably ascertains pathogenicity through similar mechanisms as caused due to p.Pro596His variants [14,15].

Two carriers were found to harbor variants in *FAT1*; AT2176 had two Het variants (p.Thr874Met and p.Phe3590Leu) while AT2029 had one Het variant (p.Asp1554Asn). AT2176 with AO of 24 years presented with gait, intention tremor, vision impairment, dysarthria, head tremor, nystagmus, mild bradykinesia, and psychiatric symptoms and carried a differential clinical diagnosis of Friedreich ataxia (FRDA)-like or a mitochondrial disorder. AT2029 on the other hand, had an AO of 50.5 years and presented with cerebellar signs, gait, intention tremor and dysarthria only. Autosomal dominant (AD) mode of inheritance has been reported for *FAT1 in* cerebellar ataxias with an AO ranging between 10-70 years and variable clinical presentation[16], while reported cases of childhood-onset (<10 years) patients with episodic ataxia have p.Asp1930His variant [16].

Of the two variants identified in AT2176, p.Phe3590Leu was found to have 44 Het in the gnomAD database and has been classified as a VUS using the ACMG-AMP criteria. Protein modelling studies of these variants have implicated their role in the alteration of the protein structure (Supplementary Fig. S2 and S3). Due to the unavailability of the parent's sample for AT2176, we were unable to test the two variants for cis or trans (CHet) heterozygous state. AT2176 also reported a pathogenic variant, W22R [17] in Het state in autosomal recessive (AR) inherited gene, *NDUFB3* as an incidental finding (Supplementary file 2). We conclusively designated the p.Thr874Met variant for the causative genetic defect in our indexed patient AT2176. Functional analysis of the same further explained the disease severity and early age of onset in the patient AT2176.

Further, we observed variant p.Ser399Leu in *KCNC3* (SCA13, 605259) and a previously reported variant p.Gly230Val [18] in *ELOVL5* (SCA38, 615957) in sporadic patients AT2216 and AT1889 respectively. An asymptomatic younger family member of the AT2216 was also found to harbor the p.Ser399Leu variant. Variant p.Gly230Val in *ELOVL5* was also observed in asymptomatic offspring of indexed patients. The above findings highlight the importance of revaluation of asymptomatic family members, which could help in timely assessment of the disease prognosis.

Amongst all the identified CA subtypes variants, those in *SETX* were found to be the most common. Three SPEOCA families had compound heterozygous (CHet) variants in *SETX* representing SCAR1 phenotype. One ADCA family had the Het variant. Out of the seven variants observed in *SETX*, 3 Het missense variants i.e. p.Tyr324Cys, p.Ser2158Asn and p.Thr2373Ala are novel variants and segregated in CHet fashion.

Four families characterized with three HSP subtypes exhibited an overlapping phenotypic representation with CAs and therefore, broadened the findings of these two disorders [19]. We adopted a composite approach of inclusion of HSP genes in the ataxia gene panel, which resolved these uncharacterized patients with homozygous (Hmz) or CHet variants in *SPG7*, *GBA2,* and *CAPN1* in autosomal recessive (AR) mode.

In the present study, we further observed variants in 10 genes, which have been reported in other neurological disorders. Besides ataxia, these patients also presented other overlapping clinical features. The aforementioned neurological genes observed in this study have been previously reported in CA cohorts from other populations as well, indicating the pleiotropic effect of these genes, thereby serving as rare loci for ataxia [20].

We identified three AR genes, *CEP290*, *RELN,* and *NPC1*, which have been assigned as subtypes for autosomal recessive cerebellar ataxias[21]. The Hmz variant p.Leu1637Pro observed in *CEP290*, lies at the coiled-coil hotspot domain for Joubert syndrome 5 (610188)[22]. We also identified Hmz missense variant, p.Asp1943His which lies in the exon 39 in *RELN* (Lissencephaly 2, 257320) and an already reported splice site variant IVS37AS, G-A, -1 which results in deletion of exon 36 [23]. Identified variant p.Asp1943His lies in the reelin repeat 4 sub-repeat 2 (residue 1800-1949) region in *RELN* within a galactose-binding-like domain, suggesting that the observed variant is present at the hot spot site for lissencephaly 2, which is presented with ataxia.

PSEN1 Het variant p.Asn190Asp is a rare missense variant that lies in exon 7 within topological domain 4. Het variants p.Tyr115Cys, p.Glu280Gly, p.Ser290Cys and p.Met233Val have been reported to have cerebellar presentation in *PSEN1* related disease (Alzheimer disease, 607882) in an AD mode while p.Met233Val carriers are known to have spinocerebellar-like phenotype [24,25]. The spectrum of *PSEN1* variants across the exon 7, its pleiotropic nature and the damaging effect of p.Asn190Asp variant highlights its potential as a likely pathogenic diagnostic variant.

Family based design identified recently reported CAs subtype and new gene

After screening and searching variants in ataxia gene panel in index patients of S-WES design, we followed F-WES design in the remaining seven uncharacterized families (Supplementary Fig. S4). Following segregation and other pathogenicity assessment, we identified genetic determinants in four of these families.

In one SPEOCA patient, we observed CHet missense variants (p.Ser662Phe and p.Leu1383Met) in *SPTB* that were segregated in parents, absent in ethnically matched control and predicted to be protein damaging (Figure 2A-2D). *SPTB* has been reported in hereditary spherocytosis (HS) [26]. A case report by McCann & Jacob (1976) had reported two different patients with spinal cord disease who presented with loss of balance, spasticity, intention tremor, unsteady gait, and also had HS [27]. We further tested the HS phenotype in the indexed patient through the osmotic fragility test and observed an altered osmotic

fragility level (Figure 2B). This confirmed that the *SPTB* carrier has a mild HS phenotype in addition to the CA symptoms. Recently, Het variants in *SPTB* have been reported in autism spectrum disorder [28,29] and CHet variants in amyotrophic lateral sclerosis [30]. Combining the above evidence indicated *SPTB* as a candidate gene for CAs with HS phenotype.

Two of the ADCA characterized families were found to have variants in *MME* and *CAPN1* (Supplementary Fig. S5A-S5G) reported for SCA43 and Spastic Paraplegia 76 (SPG76, 616907) respectively[31,32]. MME is a neprilysin membrane metalloendopeptidase containing a transmembrane domain and an extracellular region composed of two peptidase M13 domains connected by a linker region. Observed variant p.Arg374Lys in AT2832 family and reported variant, p.Cys143Tyr[32] in *MME* lie within the peptidase M13-N domain, a hot spot region. Variant p.Arg374Lys was present in three affected family members, which underscores its role as a likely pathogenic variant.

In AT1796 family, we identified a likely pathogenic variant, p.Gly220Arg in *CAPN1* (SPG76, 616907). Three affected siblings in this family exhibited cerebellar with pyramidal and extra extrapyramidal clinical involvement suggesting spastic ataxia phenotype. CAPN1 is a calcium-regulated non-lysosomal thiol-protease, which catalyses limited proteolysis of substrates, involved in cytoskeletal remodeling and signal transduction. After calcium ion binding, CAPN1 translocates to the plasma membrane[33,34]. Variant p.Gly220Arg lies at the calpain catalytic domain and is thought to affect the calcium ion binding and subsequent processes. Variant p.Gly220Arg was predicted to be a damaging through in-silico analysis and protein modeling and segregated in three affected siblings suggested causative genetic defect in AT1796 family.

In one of the ADCA family, we observed a novel variant, p.Val2260Asp in *FAT2* (SCA44, 617769). FAT2, an atypical protocadherin2 has 34 cadherin repeats, two EGF motifs, a laminin A-G motif, a transmembrane domain and a cytoplasmic domain. The putative variant p.Val2260Asp in *FAT2* maps within the cadherin19 repeat region suggesting that perturbations of *FAT2* in its cadherin related functions (Supplementary Fig. S5H-S5K) [35,36]. Variant p.Val2260Asp is protein damaging, absent in unaffected mother and AT2221 symptoms clinically correlated with reported patients[16] of *FAT2*. This confounds *FAT2* as a causative genetic aberration in AT2221.

Variants of Uncertain significance and incidental findings

In two of the families, we observed two VUS variants i.e. p.Asn1759Ser in *SPTBN2* (SCA5, 600224) and p.Arg1591Trp in *KIF26B.* SPEOCA patient AT2833, harbor p.Asn1759Ser variant, which is localized at spectrin 14 repeat, has a gnomAD[37] frequency of less than 0.1% (0.0001276) though has more than eight heterozygotes present within the database, which is higher for AD inheritance. Phenotypic representation, cerebellar and vermian atrophy in AT2833 brain MRI overlaps with that reported in SCA5 patients. Re-evaluation of the patient's family history, subsequent analysis of variant segregation, and functional assessment of the variant would ascertained the pathogenicity of the p.Asn1759Ser VUS variant.

Another rare and damaging variant i.e. p.Arg1591Trp in *KIF26B was* observed in SPEOCA patient AT1893, with cerebellar signs and seizures. A VUS variant, p.Asp1904Asn in *KIF26B* has been recently reported in cerebellar ataxia with spasticity presentation[16]. Variant p.Arg1591Trp has less than five heterozygotes present within the gnomAD, however, no missense pathogenic/likely pathogenic variants have been reported so far, in the variants localized domain. We classified p.Arg1591Trp variant as a VUS due to lack of any evidence supporting its pathogenicity such as its presence in other positive families or any functional validation.

Incidentally, we found 30 additional variants in the ataxia gene panel, of which 26 were present in the WES, characterized cases and four were observed in the remaining unsolved cases (Supplementary file 2).

Functional insight in cell line based model

To test the efficacy of Lymphoblastoid cell line (LCLs) as a means for functional validation of CAs causing variants in a rapid and effective manner, we performed assays for pathways associated with neurodegeneration. We inferred the pathogenicity of *KCNC3* variant p.Ser399Leu in patient-derived LCLs while taking a FRDA-LCLs and a healthy donor derived LCLs as assay positive and negative controls respectively. Basal Reactive Oxygen Species (ROS) levels in KCNC3-LCLs and FRDA-LCLs showed 2-fold and 1.34-fold change respectively as compared to the negative control-LCLs (Figure 3A and 3B). Induced ROS level in AT2216, suggestive of perturbation at Endoplasmic reticulum-mitochondrial interface. Mitochondrial dysfunction and calcium homeostasis could be tested in AT2216 to ascertain involvement of converging pathways as reported in other CAs and neurodegenerative disorder [38–40].

Unfolded protein response (UPR) assay, which checked using BiP expression, showed slight induction in FRDA-LCLs, while markedly induction was observed in KCNC3-LCLs. Induced expression of BiP suggests putative pathological perturbation in UPR. However, the expression level of CHOP protein found to be very low in all LCLs (Figure 3C). This could either be due to transient expression of CHOP, homeostatic maintenances of cells or involvement of other cellular processes.

Further, to elucidate the functional role of *SPTB* in a cell line model, we used the SKNSH cell line, as the expression of SPTB RNA and protein is very low in LCLs (Supplementary Fig. S6). We prepared a transient knockdown model using siRNA approach, designed at variant location. An average of 40-50% knockdown efficiency was achieved in SPTB siRNA treated cells when compared to the scrambled siRNA treated cells (Supplementary Fig. S7). With the same knockdown condition, cell viability in SPTB siRNA treated cells reduced to 82% as compared to the scrambled siRNA treated control cells (Figure 3D). This suggests that *SPTB* may have a functional role in human neuronal cells. Further variant specific functional validation in a disease model is warranted to elaborate its pathophysiology in CAs.

Protein modelling

Molecular Dynamics (MD) simulation of SPTB, FAT1, FAT2 protein modeled domain, and protein domain of CAPN1 exhibited altered Root-mean-square deviation (RMSD) and alpha-carbon fluctuations (Supplementary Fig. S2).

Of the CHet variants p.Ser662Phe and p.Leu1383Met in SPTB, residue 662 lies within the fourth spectrin repeat while residue 1383 is present within the spacer region between the 10th and 11th spectrin repeats, close to the interacting site of ankyrin R and postsynaptic density (PSDs) scaffolding proteins. MD simulation of CHet variants of SPTB pointed towards opening of the loop in the vicinity of residue 662, while the helix region near residue 1383 changes to a coil conformation. These variants may therefore, possibly affect the binding of SPTB with its interactants (Figure 2D and Supplementary Fig. S2 and S3).

A simulation study of two variants in FAT1 (p.Thr874Met and p.Phe3590Leu) observed in AT2176 showed that the structure of variant protein is more compact than the wild type protein. The simulation results further indicated that the coil region localized in residue 3590 which forms a small helix in the wild type transformed to a beta-sheet in the variant (Supplementary Fig. S3).

Docking of 1,2-Ethanediol (EDO) in MME variant (p.Arg374Lys) localized site elucidated an altered orientation of the binding site, thereby affecting the binding of EDO (Supplementary Fig. S5D). MD simulation of FAT2 domains showed that in the wild type residue, Val2260 adopted a sheet-coil-sheet conformation while a more open structure was found in the variant residue, Asp2260 and extended beta-sheets around the variant localized region remain unchanged. Simulation of p.Gly220Arg variant containing region in CAPN1 suggests that Arg220 residue compacts the overall structure of the protein and possibly affects the binding of calcium ion at its binding site (Supplementary Fig. S2 and S3).

Discussion

In the present study, we identified a diagnostic utility variant in 50% (25/50) of the selected families. Definitive/probable diagnostic yield with a clinically relevant pathogenic/likely pathogenic variant, reported by us, is much higher (64%) than other recent studies that have extensively utilized genetic and/or functional approach [16]. Our experimental design enabled us to unveil many rare CA subtypes and ataxias presented with other neurological disorders within our cohort. The deviation from the reported phenotype in our selected patients highlighted the pleiotropic effect of genes (such as *GPR88*, *PSEN1*, and other neurological disorders genes from our study) linked to neurodegeneration. Observed known variants in ataxia reported genes in our population signifies founder events of the past, while identification of novel variants also suggests population-specific disease variants. Variants in *SETX* are the most common among characterized families. We obtained the highest diagnostic yield in the SPEOCA group of patients (Table1). `

We propose *SPTB* as a new candidate gene for CAs since it has not been reported so far to the best of our knowledge for involvement in any cerebellar phenotype in any model system. Along with neurological symptoms, we observed mild HS phenotype in AT1929 suggesting that the variants in *SPTB* could lead to

co-occurrence of HS and CAs phenotype. *SPTB* shows an abundant expression in the PSDs in Purkinje neuron[41], plays an important role in synaptic plasticity along with other PSDs and is involved in *NCAM* mediated neurite outgrowth[42]**.** Knockdown of the β1 spectrin in mice resulted in an increase in the number of perforated neurons and AMPA receptor (AMPAR) endocytosis, leading to an abnormal synaptic activity [43]**.** Overexpression of the actin-binding domain of SPTB leads to an increased number of dendritic spine, size of PSDs, and less number of functional synapses[44]. In spectrin protein families, multiple isoforms (two alpha spectrin and five-beta spectrin) have been reported with diverse cellular localization, thus indicating their variable functions. Variants in different spectrins have been implicated in both neurological and erythrocytic disorders [42,45–47]. In the beta spectrin family, *SPTBN2* variant has been associated with SCA5 and SCAR14[42,45] while variant in *SPTBN4* has been reported to cause congenital myopathy with neuropathy and deafness[47]. The cumulative evidence of reported neuronal function of SPTB and its genetic aberrations along with other spectrins leading to a neurological manifestation underscores the relevance of *SPTB* as a new candidate gene for CAs.

So far, the genetic diagnosis and mechanistic understanding of heterogeneous neurological disorders is limited by numerous constraints. The research domain in CAs diagnostics is largely restricted by the unavailability of family members' samples for segregation analysis or that of other affected families with the same gene. The reasons for not obtaining yield in all the cases with WES signifies its limitation to sequence the GC rich regions and also the repetitive elements, while WES in our study was based on short-read sequences, however long read sequencing approach in future may deal with the better outcomes. This also necessitates the application of a better phenomics-based approach using comprehensive variants annotation and the development of rapid functional validation assays for ascertaining the pathogenicity of the novel and VUS variants.

Conclusion

Our study design represents a valuable, focused, and clinically translatable approach to delineate the genetic etiology of CA patients in a genetically and ethnically diverse Indian population. The proposed approach emphasises and remarks on the need for a population-specific gene-variant panel for rapid and cost-effective genetic diagnosis of CAs. The integrated WES and functional approach is expected to increase the clinical yield of diagnostics over time allowing for timely and well-informed clinical decisions and patient management.

Material And Methods

Study design

Our ongoing efforts of genetic investigations of cerebellar ataxia patients of our cohort (~5000 until date since 1997 inclusive of unpublished observations) utilize screening of common repeat-associated SCAs (SCA1, SCA2, SCA3, and SCA12) and FRDA and able to diagnosed 30-40% cases. To characterize the remaining unsolved cases, we adopted the WES approach. This study was an effort to know the efficacy

of WES in diagnostic yield in our heterogeneous population and the frequent, rare hereditary ataxia subtypes with their variant spectrum that would help in strategizing rapid genetic diagnosis of unsolved cases.

Our selected study of participants comprised of 101 subjects including probands and other affected, unaffected and asymptomatic family members from 50 families. We have categorized families into two groups based on family history, familial and sporadic. In the familial group, we have selected ADCA families. In sporadic family groups, we have subcategorised the families based on their AO. Families with probands AO is ≤ 40 years termed as SPEOCA and >40 years defined as SPLOCA.

Besides, 11 ethnically matched controls' with an age range of 65-90 years were also included for WES (variant prioritization). DNA samples of 257 healthy individuals of age >40 years were used for frequency estimation of the prioritized variants in the background population and performed genotyping of variants using Sequenom iPLEX Gold Massarray platform[9]. For segregation analysis, we have used samples of informative family members of 15 families.

Whole exome sequencing

In the first phase of the study, 51 patients of 50 families (14 ADCA, 22 SPEOCA, and 14 SPLOCA) from our patient cohort including 11 ethnically matched controls were selected for S-WES (Supplementary Table S5). In the second phase, seven remaining unsolved family members' samples (14 individuals, ~2-3 per family) were taken for WES in family-based design (F-WES) (Supplementary Table S6).

WES was performed using TruSeq™ Exome Enrichment and Nextera exome capture protocol (Illumina Inc., San Diego, California) followed by 100 bp paired end sequencing on Hiseq2000 [9]. On average, a total of 6GB raw data as generated for each sample was processed using the method described earlier[9,10]. Briefly, Trimming and filtering of bad quality data was performed with Trimmomatic [48].

BWA and Stampy were used for reference alignment and PCR duplicates were marked using Picard Markduplicates (http://picard.sourceforge.net/), then scores were recalibrated and realigned using Genotype Analysis toolkit (GATK) from http://www.broadinstitute.org/gatk .

Final processed *.bam file of each group samples (ADCA, SPEOCA, and SPLOCA) and 11 healthy controls samples were used for combined variant calling. Unified Genotyper module of GATK(http://www.broadinstitute.org/gatk) used for variant calling and was annotated using SNPeffect (http://snpeff.sourceforge.net) and ANNOVAR[49] tools. For WES data reanalysis, VCF file is re-annotated using updated pipelines.

Ataxia gene panel cataloguing

Phenotype-derived gene panel approach was utilized for preparing ataxia associated gene panel. Genes (N=452) of cerebellar ataxia subtypes and ataxia presented other neurological disorders extracted from

OMIM [2], HGMD[50], PanelAPP, and web resources for neuromuscular disorders (https://neuromuscular.wustl.edu/ataxia/domatax.html). Gene names curated using HGNC multi-symbol checker (Supplementary file 3).

Variant Filtering, validation, and classification

To identify molecular aberrations in rapid and focused ways, we first screened variants of our enlisted ataxia gene panel in selected S-WES cases. Further, we have filtered variants from S-WES and F-WES data and searched for atypical loci and candidate gene variants. Routinely used filtering steps for rare disorders were followed wherein protein disruptive variants (missense, nonsense, splice site, frameshift) that are novel or rare (frequency ≤0.1%) in population database, 1000GP and gnomAD [37,51] selected. Variants that were present in WES sequenced 11 ethnically control samples (ADCA-Het and Hmz variants; SPEOCA and SPLOCA- Hmz variants) were filtered out.

We applied an AD or autosomal recessive (AR) disease model based on the inheritance pattern and/or age at onset (Detailed in supplementary material and methods).

Genotype-phenotype correlation performed using reported and presented clinical features. Technical validation of selected variants performed through Sanger sequencing using designed primers pairs (Supplementary Table S7) and analytical validation of selected variants done by Sequenom, iPLEX Gold Mass array technology[9]. Segregation analysis of selected variants performed in informative family member's samples using Sanger sequencing.

ACMG-AMP guidelines [11] were followed for variants class and interpreted using VarSome (automated)[12] and InterVar (user adjusted)[13] tools. (Detailed in Supplementary materials and methods). Pathogenic and/or likely pathogenic variants with technical and/or analytical validation considered for definitive diagnosis otherwise defined as a probable diagnosis.

Protein modeling

Homology modeling was done for the domain of interest of proteins SPTB, FAT1, and FAT2 using I-TASSER [52] (Supplementary Table S8). Molecular Dynamics (MD) Simulation was performed using GROMACS (v5.1.2)[53] to study the effect of variants on the domain structure. Molecular docking was also performed using the AutoDock (v.4.2.6)[54] for MME taking its reported substrate EDO (Detailed in Supplementary materials and methods).

Cellular assay in the patient-derived lymphoblastoid cell line

Lymphoblastoid cell line (LCLs) of patients (SCA1, SCA2, FRDA, the novel variant carrier of *KCNC3*; SCA13 and *SPTB* carrier) and controls were established using an in-house established protocol [55]. Basal level Unfolded Protein Responses (UPR) in LCLs of SCA1, SCA2, FRDA, *KCNC3* carrier, and *SPTB* carrier were evaluated using BiP and CHOP after 48 hours' culture and positive control was generated by treating

control LCL with 4µg/ml tunicamycin for 24 hours. Reactive oxygen species (ROS) at basal levels evaluated in LCLs (*KCNC3* carrier, FRDA, and control) after 24 hours of culture using 5µM CM-H_2DCFDA dye and analysed through flow cytometry. Control LCLs treated with tert-butyl hydroperoxide (t-BHP) and used as a positive control (Supplementary methods and materials).

Cell viability assay in SPTB knockdown cell lines

Knockdown of SPTB was performed in SKNSH cell line using designed siRNA (designed in the proximity of S662F region, sense; 5'-CAUAGUCCAGGGAAGAAUAdTdT-3' and antisense; 5'-UAUUCUUCCCUGGACUAUGdTdT-3') from Sigma Aldrich. Cells were transfected with 20 nM SPTB and scrambled siRNA (Sigma) in 1:2 ratio with a vehicle, incubated for 24 hours, and tested for knockdown efficiency. Cell viability in SPTB knockdown cells checked using MTT assays. (Supplementary methods and materials).

Declarations

Ethical clearance

Institutional Ethics Committees of the participating institutes have approved this study and all recruited subjects had given their informed consent for participation.

Acknowledgements

We are grateful to all the participants of the study, the patients and their family members for the constant support and cooperation. We express our sincere thanks to Ms Shweta Sahani (Dept of Neurology, AIIMS) for her assistance in manuscript editing and proofreading. We acknowledge funding support from CSIR for this work through the project BSC0123, MLP1601 and MLP1802.

Author contributions

Conceptualization of the work and supervision: MF and MM; **Study Design:** MF, MM and RK; **Clinical Data analysis and inferences:** AKS, AG, MF; **Sample collection and validation:** VS, AKSk, IS and PS; **Experimental work and analysis:** RK; **Next generation sequencing:** RK and SSh; **Bioinformatics and computational analysis:** BU, AJ, AK and MF; **Manuscript writing and editing:** RK, and MF

Competing interests: Nothing to disclose for any authors.

References

1. Rossi, M. *et al.* Autosomal dominant cerebellar ataxias: a systematic review of clinical features. *Eur. J. Neurol.* **21**, 607–615 (2014).

2. Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F. & Hamosh, A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* **43**, D789–D798 (2015).

3. Landrum, M. J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44**, D862-868 (2016).

4. Köhler, S. *et al.* Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Res.* **47**, D1018–D1027 (2019).

5. Ngo, K. J. *et al.* A diagnostic ceiling for exome sequencing in cerebellar ataxia and related neurological disorders. *Hum. Mutat.* **41**, 487–501 (2020).

6. Yahia, A. & Stevanin, G. The History of Gene Hunting in Hereditary Spinocerebellar Degeneration: Lessons From the Past and Future Perspectives. *Front. Genet.* **12**, (2021).

7. Sun, Y. *et al.* Autosomal recessive spinocerebellar ataxia 7 (SCAR7) is caused by variants in TPP1, the gene involved in classic late-infantile neuronal ceroid lipofuscinosis 2 disease (CLN2 disease). *Hum. Mutat.* **34**, 706–713 (2013).

8. Coutelier, M. *et al.* Efficacy of Exome-Targeted Capture Sequencing to Detect Mutations in Known Cerebellar Ataxia Genes. *JAMA Neurol.* **75**, 591–599 (2018).

9. Shakya, S. *et al.* Whole exome and targeted gene sequencing to detect pathogenic recessive variants in early onset cerebellar ataxia. *Clin. Genet.* **96**, 566–574 (2019).

10. Faruq, M. *et al.* Novel mutations in typical and atypical genetic loci through exome sequencing in autosomal recessive cerebellar ataxia families. *Clin. Genet.* **86**, 335–341 (2014).

11. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **17**, 405–424 (2015).

12. Kopanos, C. *et al.* VarSome: the human genomic variant search engine. *Bioinformatics* **35**, 1978–1980 (2019).

13. Li, Q. & Wang, K. InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. *Am. J. Hum. Genet.* **100**, 267–280 (2017).

14. Hekman, K. E. *et al.* A conserved eEF2 coding variant in SCA26 leads to loss of translational fidelity and increased susceptibility to proteostatic insult. *Hum. Mol. Genet.* **21**, 5472–5483 (2012).

15. Hizli, A. A. *et al.* Phosphorylation of Eukaryotic Elongation Factor 2 (eEF2) by Cyclin A–Cyclin-Dependent Kinase 2 Regulates Its Inhibition by eEF2 Kinase. *Mol. Cell. Biol.* **33**, 596–604 (2013).

16. Nibbeling, E. A. R. *et al.* Exome sequencing and network analysis identifies shared mechanisms underlying spinocerebellar ataxia. *Brain J. Neurol.* **140**, 2860–2878 (2017).

17. Calvo, S. E. *et al.* Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing. *Sci. Transl. Med.* **4**, 118ra10 (2012).

18. Di Gregorio, E. *et al.* ELOVL5 mutations cause spinocerebellar ataxia 38. *Am. J. Hum. Genet.* **95**, 209–217 (2014).

19. Lambe, J., Monaghan, B., Munteanu, T. & Redmond, J. CAPN1 mutations broadening the hereditary spastic paraplegia/spinocerebellar ataxia phenotype. *Pract. Neurol.* **18**, 369–372 (2018).

20. Fogel, B. L. *et al.* Exome sequencing in the clinical diagnosis of sporadic or familial cerebellar ataxia. *JAMA Neurol.* **71**, 1237–1246 (2014).

21. Rossi, M. *et al.* The genetic nomenclature of recessive cerebellar ataxias. *Mov. Disord. Off. J. Mov. Disord. Soc.* **33**, 1056–1076 (2018).

22. Coppieters, F., Lefever, S., Leroy, B. P. & De Baere, E. CEP290, a gene with many faces: mutation overview and presentation of CEP290base. *Hum. Mutat.* **31**, 1097–1108 (2010).

23. Hong, S. E. *et al.* Autosomal recessive lissencephaly with cerebellar hypoplasia is associated with human RELN mutations. *Nat. Genet.* **26**, 93–96 (2000).

24. Ns, R. *et al.* Clinical phenotype and genetic associations in autosomal dominant familial Alzheimer's disease: a case series. *The Lancet. Neurology* https://pubmed.ncbi.nlm.nih.gov/27777022/ (2016).

25. Y, S., I, K. & S, I. Spinocerebellar Ataxia-Like Presentation of the M233V PSEN1 Mutation. *Cerebellum (London, England)* https://pubmed.ncbi.nlm.nih.gov/32594361/ (2020).

26. Becker, P. S., Tse, W. T., Lux, S. E. & Forget, B. G. Beta spectrin kissimmee: a spectrin variant associated with autosomal dominant hereditary spherocytosis and defective binding to protein 4.1. *J. Clin. Invest.* **92**, 612–616 (1993).

27. McCann, S. R. & Jacob, H. S. Spinal cord disease in hereditary spherocytosis: report of two cases with a hypothesized common mechanism for neurologic and red cell abnormalities. *Blood* **48**, 259–263 (1976).

28. Lim, E. T. *et al.* Rates, distribution and implications of postzygotic mosaic mutations in autism spectrum disorder. *Nat. Neurosci.* **20**, 1217–1224 (2017).

29. Takata, A. *et al.* Integrative Analyses of De Novo Mutations Provide Deeper Biological Insights into Autism Spectrum Disorder. *Cell Rep.* **22**, 734–747 (2018).

30. Steinberg, K. M., Yu, B., Koboldt, D. C., Mardis, E. R. & Pamphlett, R. Exome sequencing of case-unaffected-parents trios reveals recessive and de novo genetic variants in sporadic ALS. *Sci. Rep.* **5**, 9124 (2015).

31. Gan-Or, Z. *et al.* Mutations in CAPN1 Cause Autosomal-Recessive Hereditary Spastic Paraplegia. *Am. J. Hum. Genet.* **98**, 1038–1046 (2016).

32. Depondt, C. *et al.* MME mutation in dominant spinocerebellar ataxia with neuropathy (SCA43). *Neurol. Genet.* **2**, (2016).

33. Sorimachi, H. *et al.* A novel member of the calcium-dependent cysteine protease family. *Biol. Chem. Hoppe. Seyler* **371 Suppl**, 171–176 (1990).

34. Hsu, C.-Y. *et al.* Deimination of human filaggrin-2 promotes its proteolysis by calpain 1. *J. Biol. Chem.* **286**, 23222–23233 (2011).

35. Angst, B. D., Marcozzi, C. & Magee, A. I. The cadherin superfamily: diversity in form and function. *J. Cell Sci.* **114**, 629–641 (2001).

36. Matsui, S. *et al.* Knockdown of Fat2 by siRNA inhibits the migration of human squamous carcinoma cells. *J. Dermatol. Sci.* **51**, 207–210 (2008).

37. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *bioRxiv* 531210 (2020) doi:10.1101/531210.

38. Bezprozvanny, I. Calcium signaling and neurodegenerative diseases. *Trends Mol. Med.* **15**, 89–100 (2009).

39. Kasumu, A. & Bezprozvanny, I. Deranged calcium signaling in Purkinje cells and pathogenesis in spinocerebellar ataxia 2 (SCA2) and other ataxias. *Cerebellum Lond. Engl.* **11**, 630–639 (2012).

40. Chen, X. *et al.* Deranged calcium signaling and neurodegeneration in spinocerebellar ataxia type 3. *J. Neurosci. Off. J. Soc. Neurosci.* **28**, 12713–12724 (2008).

41. Malchiodi-Albedi, F., Ceccarini, M., Winkelmann, J. C., Morrow, J. S. & Petrucci, T. C. The 270 kDa splice variant of erythrocyte beta-spectrin (beta I sigma 2) segregates in vivo and in vitro to specific domains of cerebellar neurons. *J. Cell Sci.* **106**, 67–78 (1993).

42. Lise, S. *et al.* Recessive mutations in SPTBN2 implicate β-III spectrin in both cognitive and motor development. *PLoS Genet.* **8**, e1003074 (2012).

43. Puchkov, D., Leshchyns'ka, I., Nikonenko, A. G., Schachner, M. & Sytnyk, V. NCAM/Spectrin Complex Disassembly Results in PSD Perforation and Postsynaptic Endocytic Zone Formation. *Cereb. Cortex* **21**, 2217–2232 (2011).

44. Nestor, M. W., Cai, X., Stone, M. R., Bloch, R. J. & Thompson, S. M. The Actin Binding Domain of βI-Spectrin Regulates the Morphological and Functional Dynamics of Dendritic Spines. *PLoS ONE* **6**, (2011).

45. Ikeda, H. *et al.* Expanded polyglutamine in the Machado-Joseph disease protein induces cell death in vitro and in vivo. *Nat. Genet.* **13**, 196–202 (1996).

46. Saitsu, H. *et al.* Dominant-negative mutations in alpha-II spectrin cause West syndrome with severe cerebral hypomyelination, spastic quadriplegia, and developmental delay. *Am. J. Hum. Genet.* **86**, 881–891 (2010).

47. Knierim, E. *et al.* A recessive mutation in beta-IV-spectrin (SPTBN4) associates with congenital myopathy, neuropathy, and central deafness. *Hum. Genet.* **136**, 903–910 (2017).

48. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinforma. Oxf. Engl.* **30**, 2114–2120 (2014).

49. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).

50. Stenson, P. D. *et al.* The Human Gene Mutation Database: 2008 update. *Genome Med.* **1**, 13 (2009).

51. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

52. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40 (2008).

53. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).

54. Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791 (2009).

55. Kumar, D. *et al.* Generation of three spinocerebellar ataxia type-12 patients derived induced pluripotent stem cell lines (IGIBi002-A, IGIBi003-A and IGIBi004-A). *Stem Cell Res.* **31**, 216–221 (2018).

Tables

Tables 1-2 are available in the Supplementary Files section.

Figures

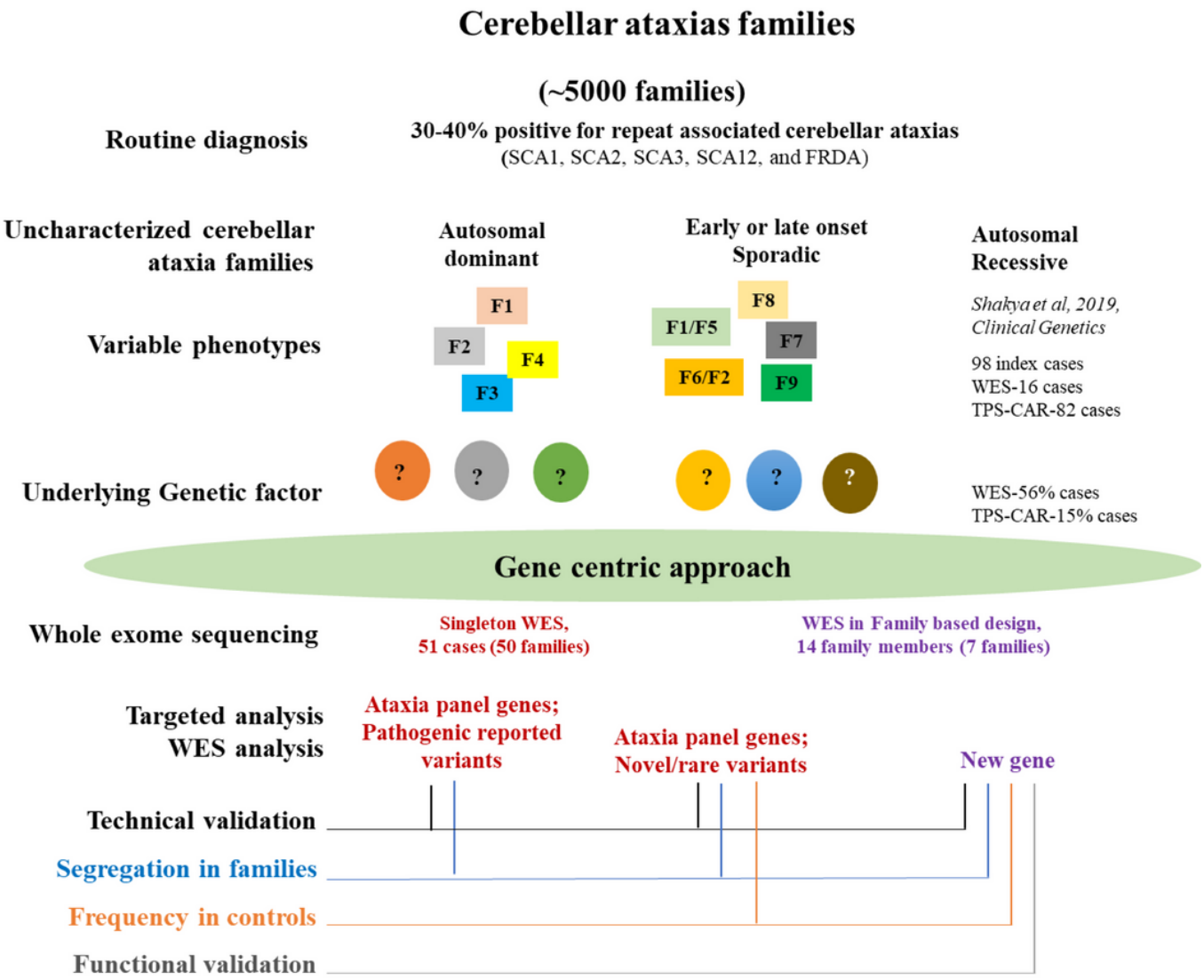

Figure 1

Technical validation, segregation analysis, frequency estimation (in 257 ethnically matched controls and in-house data), protein modelling, and genotype-phenotype correlation were performed for the prioritized variants.

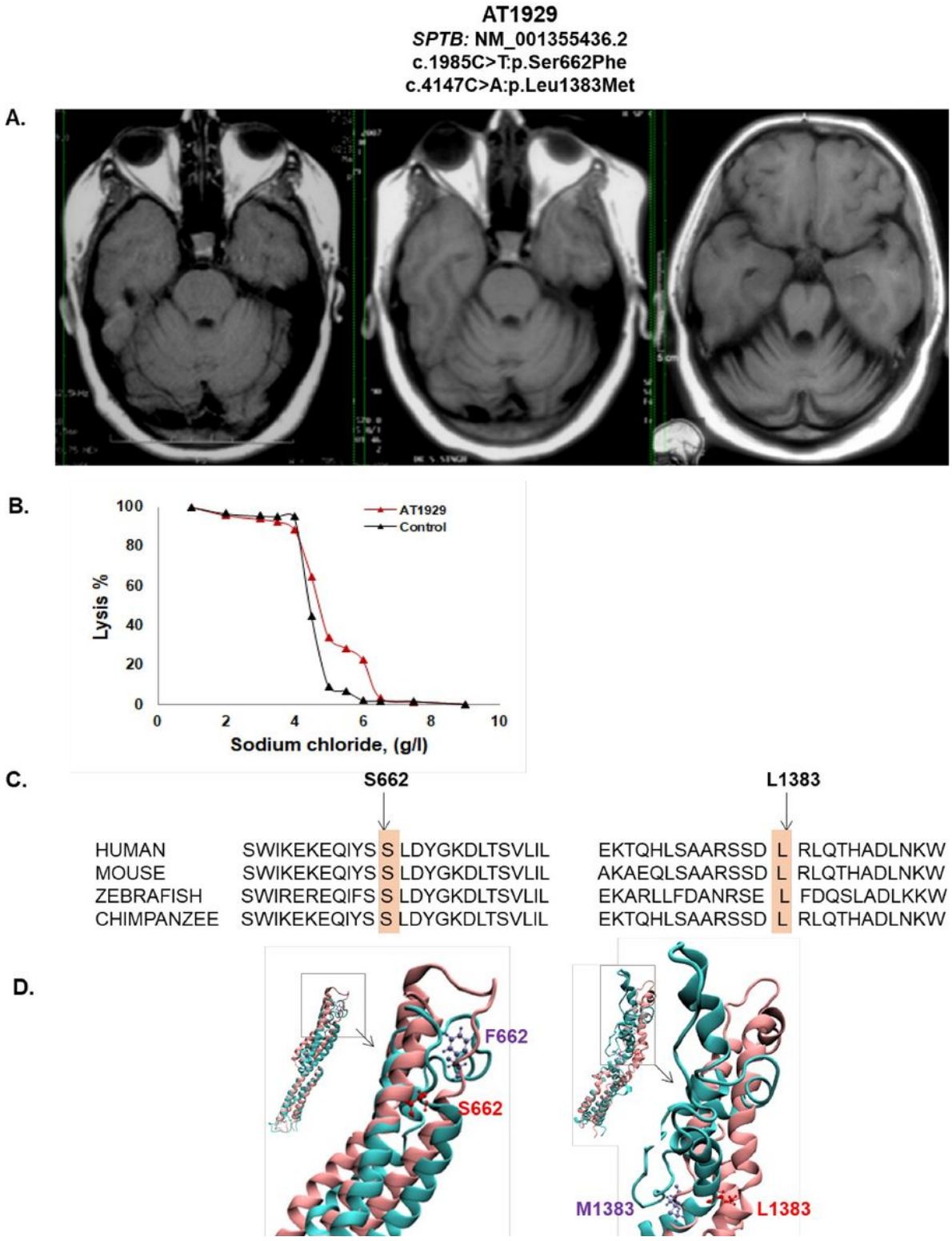

Figure 2

In one SPEOCA patient, we observed CHet missense variants (p.Ser662Phe and p.Leu1383Met) in SPTB that were segregated in parents, absent in ethnically matched control and predicted to be protein damaging .

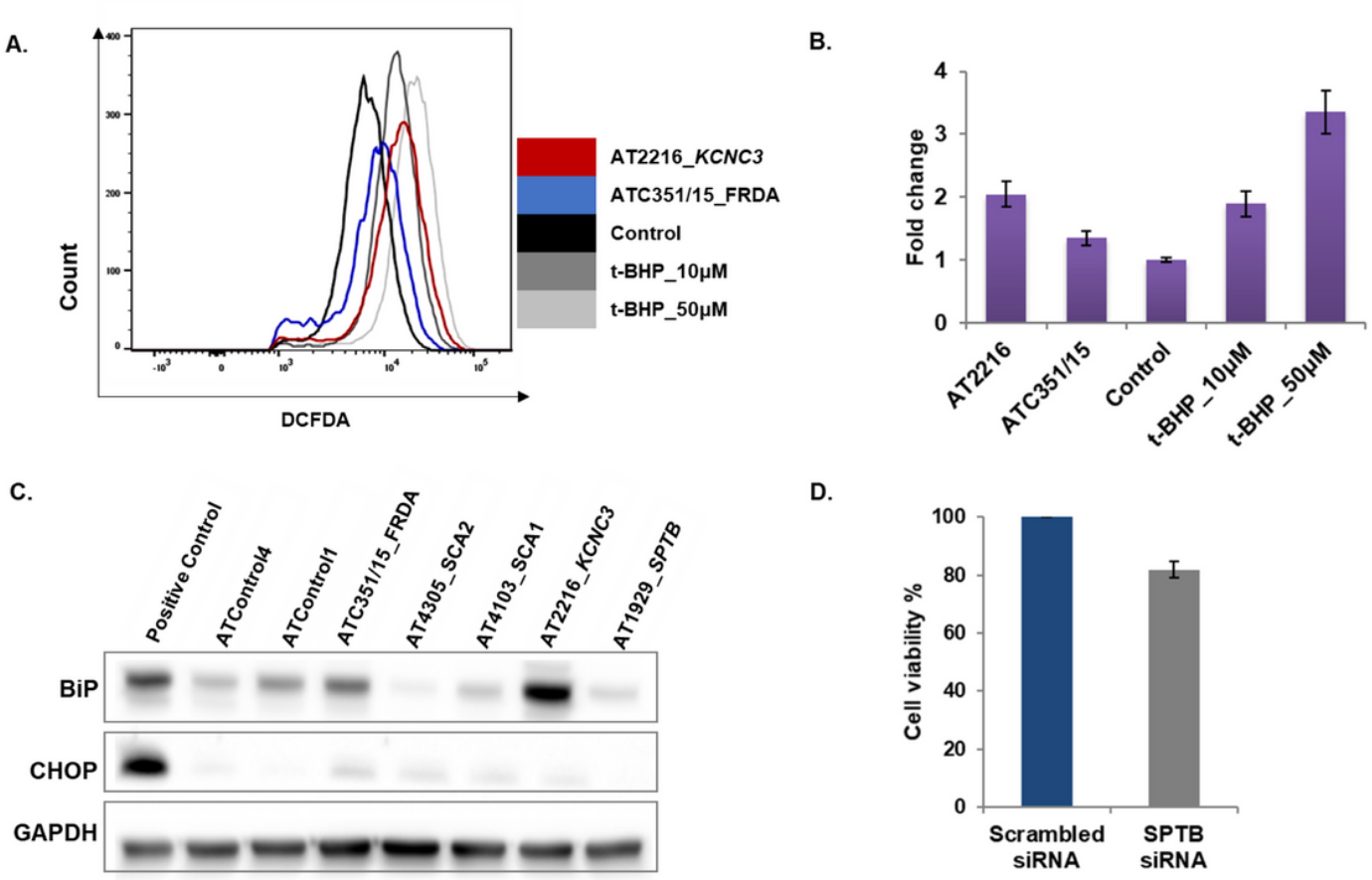

Figure 3

Basal Reactive Oxygen Species (ROS) levels in KCNC3-LCLs and FRDA-LCLs showed 2-fold and 1.34-fold change respectively as compared to the negative control-LCLs (Figure 3A and 3B). The expression level of CHOP protein found to be very low in all LCLs (Figure 3C). With the same knockdown condition, cell viability in SPTB siRNA treated cells reduced to 82% as compared to the scrambled siRNA treated control cells (Figure 3D).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Tables05072021EditMFRK.docx
- Supplementaryfile212072021Bharath.xlsx
- Supplementaryfile109062021.xlsx
- Supplementaryfile306072021.xlsx
- Supplementarymaterialsandmethods05072021.docx