UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escola Tècnica Superior d'Enginyeria Agronòmica i del Medi Natural

CHARACTERIZATION OF GENETIC BASES IN HEREDITARY ATAXIAS

Final Degree Project: Biotechnology - Academic year: 2019-2020

Author: Isabel Bustos Martínez

Tutor: Dr. Carmen Espinós Armero

Co-tutor: Dr. Elena Aller Mañas

Experimental tutor: Mª Dolores Martínez Rubio

UPV tutor: Dr. Máximo Ibo Galindo Orozco

Valencia, June 2020

Title	Characterization of genetic bases in hereditary ataxias
Author	Isabel Bustos Martínez
Tutor	Dr. Carmen Espinós Armero
Co-Tutor	Dr. Elena Aller Mañas
Experimental tutor	Mª Dolores Martínez Rubio
UPV tutor	Dr. Máximo Ibo Galindo Orozco
Location and date	Valencia, June 2020

Abstract

Hereditary ataxias constitute a heterogeneous group of movement disorders mainly characterized by progressive impairment of gait stability and coordination associated with cerebellar degeneration, sometimes affecting the speech and eye movements. According to the type of inheritance, they can be classified in ADCAs (Autosomal Dominant Cerebellar Ataxias) and ARCAs (Autosomal Recessive Cerebellar Ataxias). SCAs (Spinocerebellar Ataxias) are the most common forms of ADCAs. More than 40 subtypes have been described so far, being SCA3 the most frequent one. However, the prevalence varies among territories due to founder effects, being SCA36 the most representative in Galicia (21.3%). Regarding ARCAs, the most relevant ones are Friedreich's ataxia and ataxias with oculomotor apraxia types 1 and 2. Diagnosis is complex, especially when it comes to ADCAs due to clinical overlaps and high genetic heterogeneity. Therefore, genetic diagnosis is important to establish a definitive and accurate diagnosis.

This project includes the study of the whole exome sequencing (WES) of three families with hereditary ataxia (fATX-163, fATX-166 and fATX-167) that were negative for the prior genetic tests performed. Raw data from WES were filtered and selected following a pipeline developed in the laboratory, and a list of candidate variants was obtained. Then, validation was performed through Sanger sequencing, and in family fATX-167 the selected mutations were investigated in order to determine their possible correlation with the clinical phenotype.

SCA36 is an ADCA caused by hexanucleotide GGCCTG expansion in intron 1 of *NOP56* gene. The main objective was to design and optimize an efficient genetic test to detect expanded alleles (pathologic) through RP-PCR (Repeat Primed-PCR) and capillary electrophoresis. Furthermore, a screening was performed in a 52-patient cohort having hereditary ataxia without genetic diagnosis, including patients from families fATX-163, 166 and 167.

Nowadays, molecular diagnosis is facing challenges such as mutation's interpretation; in other words, determine whether the identified variant is the disease-causing mutation. In this FDP, the variant *POLR3A* c.3688G>A was analyzed. This variant was detected in a patient of a family suffering from spastic ataxia. *In silico* predictors revealed that this mutation, despite being located in exon 28 of the gene, could affect the splicing. Therefore, transcript analysis was performed in the proband (RNA retrotranscription, amplification of the region containing the target mutation, and subsequent electrophoresis and sequencing of the purified band).

Key words: Hereditary ataxia; SCA36 (Spinocerebellar Ataxia 36); WES (Whole Exome Sequencing); RP-PCR (Repeat Primed-PCR); Transcript analysis; *NOP56* gene; *POLR3A* gene.

Título	Caracterización de las bases genéticas de ataxias hereditarias
Autor	Isabel Bustos Martínez
Tutor	Dr. Carmen Espinós Armero
Co-Tutor	Dr. Elena Aller Mañas
Director experimental	Mª Dolores Martínez Rubio
Tutor UPV	Dr. Máximo Ibo Galindo Orozco
Lugar y fecha	Valencia, Junio 2020

Resumen

Las ataxias hereditarias constituyen un grupo heterogéneo de trastornos del movimiento caracterizados principalmente por una afectación progresiva en la estabilidad de la marcha y la coordinación asociada a degeneración del cerebelo, pudiendo haber afectación del habla y de movimientos oculares. Según el tipo de herencia se clasifican en ADCAs (*Autosomal Dominant Cerebellar Ataxias*) y ARCAs (*Autosomal Recessive Cerebellar Ataxias*). Las SCAs (*Spinocerebellar Ataxias*) constituyen la forma más común de ADCAs. Hasta la fecha se han descrito más de 40 subtipos, siendo la más frecuente la SCA3. Sin embargo, la prevalencia en cada territorio varía debido a efectos fundadores, siendo la SCA36 la más representativa en Galicia (21,3%). Respecto a las ARCAs las más relevantes son la ataxia de Friedreich y las ataxias con apraxia oculomotora tipo 1 y 2. El diagnóstico es complejo, especialmente en ADCAs debido a solapamientos clínicos, y por su acusada heterogeneidad genética. Por ello es importante el diagnóstico genético que permite establecer un diagnóstico definitivo y certero.

En el presente trabajo se incluye el estudio de la secuenciación de exomas (WES, *Whole Exome Sequencing*) de tres familias con ataxia hereditaria (fATX-163, fATX-166 y fATX-167) que han resultado negativas para todas las pruebas genéticas realizadas con anterioridad. Los datos brutos del WES se filtraron y se seleccionaron siguiendo un protocolo propio, y se obtuvo una lista de variantes candidatas. Posteriormente, se realizó su validación mediante secuenciación Sanger, y en la familia fATX-167 se estudió la segregación de las mutaciones seleccionadas para determinar una posible asociación con el fenotipo clínico.

La SCA36 es una ADCA producida por expansiones del hexanucleótido GGCCTG en el intrón 1 del gen *NOP56.* El objetivo principal fue el diseño y la optimización de un test genético eficaz para detectar alelos expandidos (patológicos) mediante RP-PCR (*Repeat Primed*-PCR) y electroforesis capilar. Además, realizamos un cribado en una cohorte de 52 pacientes con ataxia hereditaria sin diagnóstico genético establecido, incluyendo pacientes de las familias fATX-163, 166 y 167.

Uno de los desafíos a los que se enfrenta el diagnóstico molecular actualmente es la interpretación de las mutaciones; es decir, establecer que el cambio identificado es la mutación causal de la enfermedad. En este TFG, se ha investigado la variante *POLR3A* c.3688G>A identificada en un paciente de una familia con ataxia espástica. Los predictores *in silico* mostraban que este cambio, pese a localizarse en el exón 28 del gen, podría alterar el *splicing*, por lo que se llevó a cabo un análisis de transcritos en el probando (retrotranscripción del RNA, amplificación de un fragmento que contenía el cambio de interés, y posterior electroforesis y secuenciación de la banda purificada).

Palabras clave: Ataxia Hereditaria; SCA36 (*Spinocerebellar Ataxia 36*); WES (*Whole Exome Sequencing*); RP-PCR (*Repeat Primed-PCR*); Análisis de Transcritos; gen *NOP56*; gen *POLR3A.*

Acknowledgements

En primer lugar, me gustaría agradecer el apoyo y confianza recibidos por parte de mi tutora Carmen Espinós desde el primer día, a pesar de mi inexistente experiencia. Además, a Elena Aller, por su paciencia y alegría; gracias a ellas este proceso ha sido indudablemente más sencillo.

Tampoco podría olvidarme de mis compañeras de laboratorio; Paula, Ana, Amparo e Isabel y de Vincenzo, por acogerme con tanto cariño, por los innumerables consejos, por las sonrisas al llegar al laboratorio y las charlas durante las comidas. En especial quiero agradecerle a Lola su infinita paciencia y dedicación.

A mis compañeros y amigos, por afrontar las numerosas adversidades a las que nos ha sometido la universidad entre lágrimas y sobre todo humor; podría ser peor. Allá donde nos lleven los años, Benimaclet siempre será el comienzo; muchas gracias Lucía, María, Jorge, Juan, Aina, Adrián, David…

Sin duda, todos mis logros se los debo a las mujeres de mi vida, que son muchas, pero en especial a mi madre, por ser un lugar seguro al que siempre volver. A mi hermana por ser mi polo opuesto. Y a ti Fran, por enseñarme tanto, en este y en cualquier universo paralelo.

Index of content

Index of Figures

Index of Tables

Abbreviation list

AD	Autosomal Dominant
ADCA	Autosomal Dominant Cerebellar Ataxia
AFP	Alpha-Fetoprotein
Al. D	Alzheimer's Disease
AR	Autosomal Recessive
ARCA	Autosomal Recessive Cerebellar Ataxia
AOA2	Ataxia with Oculomotor Apraxia type 2
ATP	Adenosine Triphosphate
BER	Base Excision Repair
Bp	base pair
CA	Cerebellar Ataxia
cDNA	Complementary DNA
CNAG	Centre Nacional d'Anàlisi Genòmica
CNS	Central Nervous System
CNV	Copy Number Variations
DNA	Deoxyribonucleic acid
DRLPA	Dentatollubar-Pallidoluysian Atrophy
ExAC	Exome Aggregation Consortium
FA	Friedreich's Ataxia
FDP	Final Degree Project
GnomAD	The Genome Aggregation Database
HA	Hereditary ataxia
HGMD	The Human Gene Mutation Database
HPA	Human Protein Atlas
HSF	Human Splice Finder
HSP	Hereditary Spastic Paraplegia
iPSC	induced Pluripotent Stem Cells
LO	Leukodystrophy with Oligodontia
MAF	Minor Allele Frequency
MiRNA	microRNA

MRI	Magnetic Resonance Imaging
MtDNA	mitochondrial DNA
NBIA	Neurodegeneration with Brain Iron Accumulation
NCS	Neuronal Calcium Sensor
nDNA	nuclear DNA
NGS	Next Generation Sequencing
NNSplice	Neural Network Splice
OMIM	Online Mendelian Inheritance in Man.
PC	Purkinje Cell
PCR	Polymerase Chain Reaction
PD	Parkinson's Disease
PSIC	Position Specific Independent Counts
RNA	Ribonucleic Acid
ROS	Reactive Oxygen Species
RP-PCR	Repeat-Primed PCR
SCA	Spinocerebellar Ataxia
Sherloc	Semiquantitative Hierarchical Evidence-based Rules for Locus Interpretation
SIFT	Sorting Intolerant From Tolerant
SnoRNP	Small nucleolar Ribonucleoproteins
SNP	Single Nucleotide Polymorphism
SRSF2	Serine/arginine-rich splicing factor 2
STGG	Service of Translational Genetics and Genomics
TACH	Tremor-Ataxia with Central Hypomyelination
tRNA	transfer RNA
VCF	Varian Calling Format
VEP	Variant Effect Predictor
VILIP	Visin-Like Protein
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
4H syndrome	Hypomyelination, Hypodontia, Hypogonadotropic and Hypogonadism

6-FAM	6-carboxyfluorescein
8-oxoG	8-oxoguanine

1. Introduction

1.1 Hereditary ataxias

The term ataxia refers to the inability to coordinate voluntary movements leading to gait imbalance and instability usually caused by atrophy in the cerebellum or cerebellar pathways; nevertheless sometimes it involves the spinal cord. In this sense, symptoms may vary according to the damaged region; for instance atrophy in the cerebellum vermis leads to truncal and gait ataxia (Ashizawa *et al*., 2016). Other symptoms may occur as a result of cerebellar degeneration; including: incoordination of the upper limbs affecting hands and fingers, nystagmus or dysarthria among others. Ataxia is a clinical manifestation associated with a wide range of pathologies and alterations; therefore, two different groups can be distinguished: acquired and inherited ataxias (Albernaz *et al*., 2019).

Acquired ataxias can be produced by alcoholism, haemorrhage, cerebellar infarction, primary tumors and metastasis or multisystemic atrophy leading to cerebellar degeneration.

When it comes to hereditary ataxias (HA), they constitute a heterogeneous group, both clinically and genetically, of movement disorders. They are classified according to the type of inheritance; having ADCAs (Autosomal Dominant Cerebellar Ataxia), ARCAs (Autosomal Recessive Cerebellar Ataxia), X-linked and mitochondrial inherited ataxias.

ADCAs are the most common ones having a worldwide prevalence of around 1.5-4:100,000; although it differs significantly among populations and territories (Ruano *et al*., 2014). Several types can be distinguished: SCAs (Spinocerebellar Ataxia), episodic ataxias and a complex form called DRLPA (Dentatollubar-Pallidoluysian Atrophy). SCAs are the most common ones and more than 40 different subtypes have been described so far; all presenting cerebellar degeneration that cannot be differentiated with MRI (Magnetic Resonance Imaging). Diagnosis is complex due to clinical overlapping (Shakkottai *et al*., 2013); nonetheless they usually present extracerebellar symptoms that may enable to distinguish between the different subtypes. For instance, SCA7 is characterized by retinal degeneration (Jayadev *et al*., 2013). Nevertheless, molecular genetic testing becomes crucial to establish a definite diagnosis. Most prevalent SCAs are produced by polyglutamine (CAG) expansions in coding regions; causing SCA1-3, SCA6, SCA7, SCA17 and DRLPA. Consequently, insoluble protein aggregates are detected both in the cytoplasm and the nucleus, although they have not been proven to be pathological (Seidel *et al*., 2012). In addition, repeat expansions may produce genetic instability, especially during meiosis, leading to anticipation. This process consists of the abnormal expansion of the repeats in subsequent generations, presenting an earlier age of onset and more severe symptoms. Anticipation is recurrent in SCA7 (Seidel *et al*., 2012). SCA3 is the most common form worldwide; however its prevalence varies significantly among territories due to founder effects (Jayadev *et al*., 2013) as it occurs with SCA36, the most common ADCAs in the northwestern region of Spain (Garcia-Murias *et al*., 2012).

When it comes to ARCAs, the first clinical manifestations appear in the childhood or early adulthood and patients usually present with a more severe phenotype than ADCAs. The most common type is Friedreich's ataxia (FA), which is caused by trinucleotide GAA expansions in *FXN* gene (Bürk, 2017). Other common forms include ataxia with oculomotor apraxia types 1 and 2 (AOA2) and ataxia telangiectasia.

1.1.1 Spinocerebellar ataxia 36

SCA36 is a form of ADCA characterized by slowly progressive cerebellar ataxia of late onset. This form of ataxia was firstly identified in three unrelated families of Chugoku district (Japan) presenting gait and limb ataxia, dysarthria and abnormal eye movements (Kobayashi *et*

al., 2011). Furthermore, motor neurons were affected in long-term patients, leading to atrophy and fasciculation of skeletal muscles and the tongue. The involvement of motor neurons in other types of ataxia, for instance SCA1, is minimal, thus constituting a distinct trait of SCA36. MRI shows atrophy in the cerebellar vermis in the early stages, expanding to other cerebellar regions as the disease progresses. Shortly after, this pathology was also described in the northwestern region of Spain, Costa da Morte, with an additional symptom: progressive hearing loss (Garcia-Murias *et al.*, 2012).

SCA36 is produced by a hexanucleotide GGCCTG expansion in intron 1 of *NOP56* gene; located in chromosome 20. This protein is a component of the C/D small nucleolar ribonucleoprotein (snoRNP) complex which is involved in ribosomal synthesis and splicing (Yang *et al.*, 2020) (**Figure 1**). Expanded alleles (pathogenic) contain between 600 to 2500 repeats. As a result, it has been observed that RNA foci accumulate in the nucleus of lymphoblastic cell lines of SCA36 patients (gain of toxic function). Nevertheless, protein expression levels are not affected (Ikeda *et al.*, 2012). As it occurs in other forms of SCA, RNA foci capture specific RNA binding proteins; in this case, SRSF2 which is involved in splicing (Kobayashi *et al.*, 2011). In addition, the expansion leads to downregulation of MIR1292 (miRNA), located 19 bp downstream from the expansion. Several studies have associated miRNA dysregulation with neurodegenerative disorders (Hébert *et al.*, 2009). In this sense, downregulation of MIR1292 affects signal transduction to Purkinje cells (PC) leading to ataxia (Kobayashi *et al.*, 2011), although further studies are required to determine the contribution of RNA foci and MIR1292 to pathogenesis.

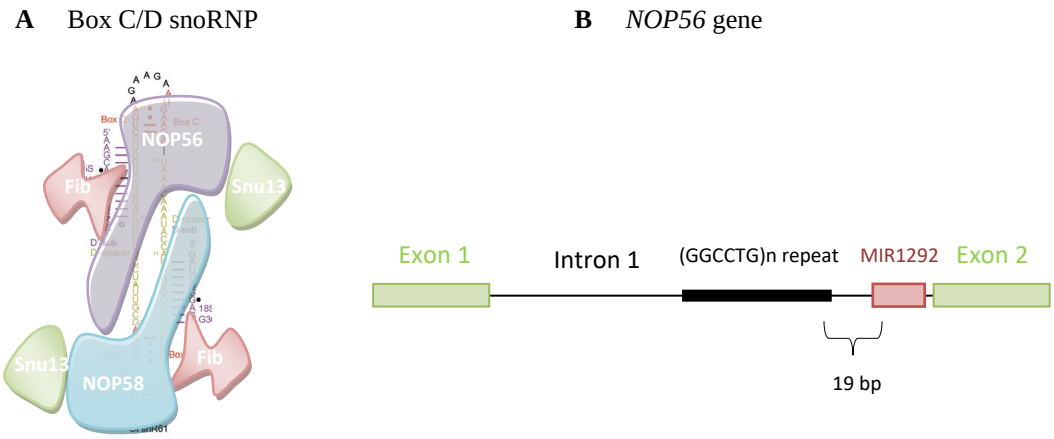

Figure 1: A. Structural representation of the box C/D snoRNP. It is composed of RNA and proteins, including NOP56, NOP58, fibrillarin and Snu13. This functional complex is able to bind to guide RNAs. **B**. Representation of intron 1 of *NOP56* gene; containing the hexanucleotide repeat and the location of MIR1292 (miRNA).

Many neurodegenerative disorders have been related with repeat expansions, however nowadays there is not any effective treatment available for none of them, including SCA36. Therefore, clinical manifestations are mitigated with physical therapy to improve gait balance, walking aids and speech therapy among others. Recent studies have focused on the reduction of RNA foci as a promising therapeutic target for SCA36. In this sense, Matsuzono *et al.* (2017) established a cellular model of SCA36 with iPSC (induced Pluripotent Stem Cells) and neuron derived iPSC from patients. Cells were treated with antisense oligonucleotides specific for *NOP56* pre-mRNA reporting a 50% decrease in the RNA-foci positive cells, thus constituting a promising therapeutic tool not only for SCA36 but for other repeat expansion diseases.

Although the molecular mechanism of SCA36 is common in both geographic areas there are some differences regarding genetics. Screening studies in large cohorts of controls suggested that non-expanded alleles in Japanese population have between 3 to 8 repeats (Kobayashi *et al.*, 2011) while in Spanish population it ranges between 5 to 14 hexanucleotide repeats, being 9 the most common allele (Garcia-Murias *et al.*, 2012). On the other hand, pathogenic alleles present between 600 to 2500 hexanucleotide repeats and have full penetrance (Arias *et al.*, 2014).

Furthermore, the presence of the expansion leads to genomic instability, especially during meiosis which may lead to the phenomena of anticipation. The abnormal expansion of the repeats in subsequent generations is more common in paternal transmission than in the maternal one; where contraction is sometimes detected.Consequently, the offspring presents a more severe form of SCA36; usually with earlier age of onset. This process has been detected in a family from Costa da Morte region (Galicia) (Garcia-Murias *et al.*, 2012). SCA36 patients in each of the geographic regions present a common haplotype of 1.8 Mb in Chugoku (Japan) and of 0.8 Mb in Costa da Morte, suggesting founder effects. Further studies are required to determine whether the Japanese and Spanish haplotype have the same origin. As a result, SCA36 is the most common form of ADCA in this Spanish region, accounting for the 21.3% of the cases. Few cases have been detected in unrelated families in the rest of Spain.

Clinical diagnosis of SCA36 involves the consideration of several factors: clinical manifestations, a family history corresponding to an autosomal dominant (AD) pattern of inheritance as well as the geographic area of the affected individuals. Nevertheless, molecular genetic testing is required in order to confirm the diagnosis. In this sense, Repeat Primed-PCR (RP-PCR) and capillary electrophoresis are routinely used to identify pathologic alleles with high efficiency due to full penetrance (Arias *et al.*, 2014). In addition, the hexanucleotide expansion is exclusively found in SCA36 patients as it has been detected neither in 234 unaffected individuals nor in 214 patients presenting spastic paraplegia in Spain (Garcia-Murias *et al.*, 2012). Other techniques such as Southern Blot can be applied to estimate the number of repeats in expanded alleles. In contrast, non-expanded alleles (up to 15 repeats) are detected with standard PCR. Alleles presenting an intermediate length (between 15 to 600 repeats) are of unknown significance and can be detected by standard PCR or RP-PCR according to their length.

1.2 Hereditary spastic paraparesias

Hereditary spastic paraparesias (HSP) are a clinically and genetically heterogeneous group of neurodegenerative and neurodevelopmental diseases, primarily characterized by lower limb spasticity (permanent muscle contraction) and weakness. HSP's clinical manifestations are produced as a result of distal axonal degeneration of the long motor corticoespinal (descending) and dorsal (ascending) tracts (Salinas *et al.*, 2008). The corticoespinal tract, one of the two pathways in which the pyramidal tract divides, is composed of efferent nerve fibers involved in motor signal transmission from the cerebral cortex to the spinal cord, ultimately controlling voluntary movements of the trunk and lower limbs (Welniarz *et al.*, 2017).

HSP are frequently classified according to the clinical phenotype, having pure and complex HSP forms. Furthermore, they can be transmitted in an AD, AR, X-linked or mitochondrial maternal pattern of inheritance.

Pure HSP presents with symptoms exclusively related with corticoespinal tract degeneration; including: spasticity, weakness and mild sensory loss in the lower limbs and hyperreflexia (S. de Souza *et al.*, 2016). Urinary dysfunction is observed in 50% of the patients. Generally, pure HSP presents an AD pattern on inheritance, being *SPAST*-associated HSP the most prevalent one, accounting for the 40% of the familial cases (Salinas *et al.*, 2008). It also constitutes the most common cause of sporadic cases.

By contrast, complex forms, normally AR types, are associated with a wide range of other clinical symptoms, both neurological and non-neurological. Neurological symptoms may include cognitive impairment (such as mental retardation or dementia), epilepsy, extrapyramidal features (like dystonia) and cerebellar dysfunction, leading to ataxia, nystagmus, dysarthria or tremor (S. de Souza *et al*., 2016). Therefore, HSP and spinocerebellar ataxias present overlapping phenotypes and genes (Synofzik *et al.*, 2017). Non-neurological manifestations comprise ophthalmological dysfunctions such as optic atrophy or retinitis pigmentosa and orthopedic disorders, for instance, scoliosis.

The age of onset is remarkably variable ranging from early childhood, especially in AR forms, to 70 years of age (Kara *et al.*, 2016). Patients present with subtle initial symptoms consisting of leg rigidity (Salinas *et al.*, 2008). AD forms typically have an age of onset in the adulthood. HSP worldwide prevalence is around 2-5:100,000 (Shribman *et al*., 2019) although it differs among territories.

As previously stated, HSP constitutes a genetically heterogeneous group of movement disorders caused by axonal degeneration of corticoespinal tract long fibers which can be up to one meter long. Due to its length, they depend intensively on mechanisms regulating axonal development and maintenance, especially in the distal axonal parts. These mechanisms comprise membrane transport, mitochondrial function, axonal transport of macromolecules and organelles, lipid metabolism and organelle biogenesis (Blackstone, 2018). Regardless of the fact that the exact pathophysiological mechanisms have not been determined yet; more than 60 genes have been related to HSP so far (Parodi *et al.*, 2018), most of them involved in the regulation of the aforementioned pathways.
For instance, mutations in *SPAST* gene generally lead to spastin inactivation, which is a hexameric microtubule serving ATP-ase protein (Solowoska *et al.*, 2015). *SPAST* encodes for two isoforms: M1 spastin, mainly located in the endoplasmic reticulum membrane, and M87 spastin located in the cytoplasm. They regulate microtubule mobility, length and distribution. Inactivating mutations in *SPAST* gene lead to abnormalities in axonal and membrane transport in the distal axon, occurring in 40-45% of AD HSP (Salinas *et al.*, 2008). Nevertheless, inactivating mutations in *SPAST* gene do not explain the whole *SPAST*-HSP phenotype, therefore it has been suggested that the mutant hexameric protein may also lead to neurotoxicity (Blackstone, 2018). Furthermore, other genes are implicated in mitochondrial pathways; such as *SPG7*, which encodes for paraplegin (Synofzik *et al.*, 2017). This protein is located in the inner mitochondrial membrane and it is implicated in ribosomal assembly and protein quality control. Mutations in this gene lead to paraplegin inactivation; which impairs ATP synthesis required for axonal transport. In addition, distal axonal swelling is detected because of oxidative stress (Salinas *et al.*, 2008). Mutations in *SPG7* account for the 5% of AR HSP forms.

Nevertheless, as previously stated, ataxia and HSP share common phenotypes and genes and they commonly co-occur, leading to ataxia-spasticity spectrum disorders (Synofzik *et al.*, 2017). In addition, proteins encoded by genes traditionally involved in ataxia and HSP interact in common cellular pathways. Therefore, these disorders have common etiology mechanisms. What is more, several genes associated with HSP are implicated in ataxia, and backwards. For instance, *SPG7* is responsible for an ARCA form (Pfeffer *et al.*, 2015). Consequently, both cerebellar and corticoespinal pathways are vulnerable to the same metabolic pathways. Based on these observations, WES (Whole Exome Sequencing) will ensure a more efficient genetic diagnosis allowing the detection of a wide range of genes implicated in both overlapping disorders.

1.2.1 *POLR3A* mutation associated to HSP and cerebellar ataxias

POLR3A and *POLR3B* genes encode for two different subunits of the DNA-directed RNA-polymerase III, RPC1 and RPC2 respectively. This polymerase is involved in the synthesis of small RNAs, specifically tRNAs and the 5S ribosomal subunit.

Traditionally, pathogenic mutations identified in either of these genes were associated with a heterogeneous group of inherited neurodegenerative disorders: so-called *POLR3*-related leukodystrophy, presenting hypomyelination as a distinguishing feature (Cayami *et al*., 2015). Patients suffering from these diseases present with severe clinical manifestations involving neurologic dysfunctions, such as cerebellar and pyramidal tracts affection leading to spasticity, tremor, cerebellar ataxia or mental retardation (Bernard *et al.,* 2011). Furthermore, non-neurological symptoms are frequently detected due to *POLR3A* and *POLR3B* ubiquitous expression; primarily consisting of dental abnormalities (oligodontia, hypodontia…), and hypogonadotropic hypogonadism, which impairs hormone production by the gonads generating endocrine dysfunctions (Wolf *et al*., 2014). The first clinical manifestations appear early in childhood.

POLR3A encodes for the largest catalytic subunit of RNA-polymerase III being part of the active centre (Dumay-Odelot *et al.,* 2010). In this sense, recessive and compound heterozygous mutations detected in this gene have been associated with different overlapping disorders of the *POLR3*-related leukodystrophy spectrum; including: TACH (Tremor-Ataxia with Central Hypomyelination), LO (Leukodystrophy with Oligodontia) or 4H syndrome (Hypomyelination, Hypodontia, Hypogonadotropic and Hypogonadism) among others. However, up to 40% of the affected individuals remain genetically undiagnosed due to phenotype overlapping and unidentified non-coding mutations (Bernard *et al*., 2011).

Several studies reported that disease-causing mutations in *POLR3A* lead to a decrease in RPC1 expression levels (Bernard *et al*., 2011); hence, affecting small RNA synthesis, including tRNAs. What is more, Scheper *et al*. (2007) observed that myelin homeostasis highly depends on tRNAs regulation, hence RPC1 decreased expression levels may generate the characteristic hypomyelination observed in *POLR3*-related leukodystrophy disorders (Minnerop *et al.,* 2017).

However, as previously mentioned, diagnosis of neurodegenerative disorders is at least a challenging procedure due to genetic and phenotypic heterogeneity and clinical overlapping between them. In this sense, Minnerop *et al*. (2017) recently identified pathogenic non-coding mutations in *POLR3A* when screening an extensive cohort of patients presenting with cerebellar ataxia (CA) and HSP. In the aforementioned gene, 3.1% of the individuals harboured mutations, of which 80% carried c.1909+22G>A mutation in at least one allele. This mutation is located 22 bp downstream of the donor site of intron 14 leading to the activation of a cryptic splice site (Minnerop *et al*., 2017). Individuals positive for this mutation present a homogeneous phenotype consisting of a combination of limb and gait ataxia, pyramidal involvement, tremor in the upper/lower limbs and in some cases dysarthria or dysphagia (Infante *et al*., 2020). The first symptoms usually appeared before the second decade of life. MRI showed hyperintensities in the superior cerebellar peduncle in 80% of affected individuals (Infante *et al.,* 2020), although hypomyelination was not observed (Minnerop *et al.,* 2017, Infante *et al*. 2020), which was the distinguishing feature in *POLR3*-leukodystrophy related disorders.

As it can be observed, mutations in *POLR3A* have been associated with a wide spectrum of neurodegenerative disorders with varying severity depending on the degree of RPC1 expression inhibition. In this case, c.1909+22G>A has been associated with HSP and CA. The mechanism is based on the insertion of 19 bp from intron 14 in the coding region causing a frame-shift mutation and a premature stop codon, eventually decreasing RPC1 expression levels (Minnerop *et al.,* 2017). Nevertheless, it leads to a milder phenotype when compared with leukodystrophy because the mutation involves an incomplete activation of the cryptic splice site; allowing for the coding of reduced RPC1 levels. Other identified mutations lead to abnormal interactions with DNA, other polymerase III subunits or transcription factors.

Finally, the identification of non-coding mutations in *POLR3A*, especially c.1909+22 G>A, has enabled to broaden the spectrum of neurodegenerative disorders related to this gene, including HSP, characterized by a milder phenotype and lack of hypomyelination.

1.3 Diagnosis of neurodegenerative diseases

When it comes to neurodegenerative disorders, the establishment of a molecular diagnosis is challenging due to genetic heterogeneity and clinical overlapping among conditions. In this sense, ataxia is not the exception; multiple genetic disorders may cause progressive ataxia and many more show ataxia as a clinical manifestation (Fogel *et al.*, 2014). Furthermore, to ensure an accurate diagnosis the etiology has to be determined, distinguishing between acquired and HAs (Jayadev *et al.*, 2013). Acquired ataxias can be originated by a wide range of non-genetic causes including: alcoholism, vitamin deficiencies or primary or metastatic tumors. Depending on the origin, several treatments may be available (Jayadev *et al.*, 2013). Once clinicians have dismissed non-genetic causes of ataxia, the establishment of an accurate diagnosis depends on exhaustive examination of the medical history, clinical manifestations, neuroimaging and family history in order to acquire a deep insight of the pathology. Nevertheless, due to clinical overlapping between different ataxia subtypes, genetic analysis is required for a definitive diagnosis. In this sense, next generation sequencing (NGS) technologies constitute a high-throughput technique allowing for the detection of candidate causative mutations with higher accuracy and reduced costs when compared with prior sequencing techniques (Galatolo *et al.*, 2018). Different approaches rely on NGS, such as gene panels, WES and WGS (Whole Genome Sequencing) (Nemeth *et al.*, 2013), which have been extensively employed in the last decades. As a result, clinical diagnosis has experienced a great breakthrough identifying many ataxia and HSP causing genes. The different NGS approaches have enhanced considerably the molecular diagnosis in hereditary ataxias. In this project, WES was used as it shows the highest efficiency when applied to a heterogeneous cohort of undiagnosed HA patients (Galatolo *et al.*, 2018). Furthermore, trio-WES analysis presents a higher effectiveness; especially in sporadic childhood cases and in AR forms (Fogel *et al*., 2014). This approach permits to identify *de novo* mutations. Nevertheless, AD forms are usually of adult-onset; therefore, the applicability of trio-WES technique is limited (Pyle *et al*., 2015).

As stated before, the most common forms of ADCAs are SCAs caused by CAG trinucleotide expansions: SCA1-3, SCA6, SCA7 and SCA17. They cannot be detected by either WES or gene panels. Consequently, the first screening consists of the detection of polyglutamine expansions which enables the diagnosis of 50-60% of ADCA (Jayadev *et al.*, 2013). In addition, patients are also tested for FA, as it is the most common form of ARCA with an estimated prevalence of 2-4:100,000 (Palau *et al.*, 2006). FA is caused by GAA trinucleotide expansions in the *FXN* gene.

Then, either of the previously mentioned NGS approaches may be used. Gene panels sequence the coding exons of a set of genes, while WES analyses the coding region of the genome (1-2% of the whole genome). Finally, WGS sequences coding and non-coding regions of the genome and allow the detection of copy number variations (CNV). Nonetheless, several algorithms have been developed lately to predict CNV from WES data, increasing the yield of this technique (Marchuk *et al*., 2018). This diagnosis procedure is represented in **Figure 2**.

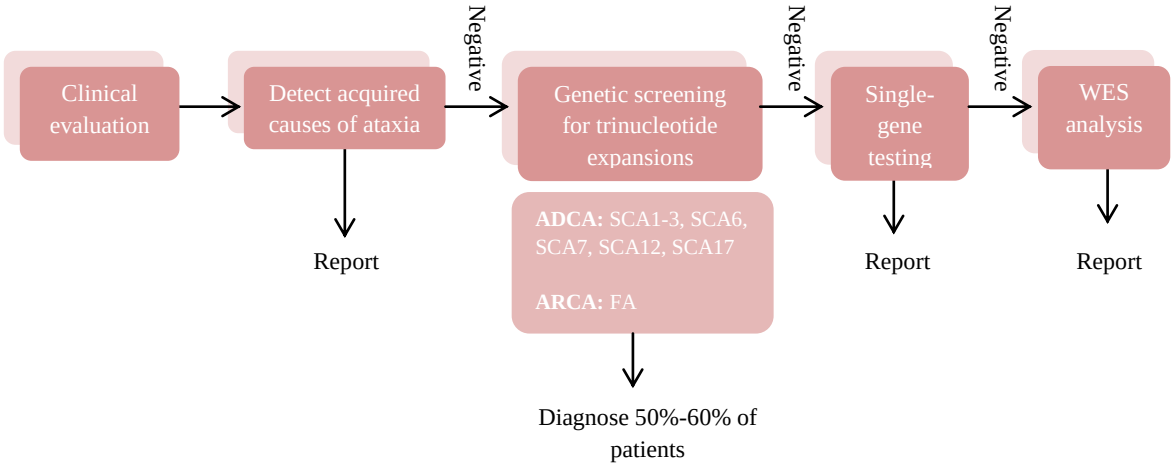

Figure 2: Flowchart of diagnosis procedure for patients presenting with progressive cerebellar ataxia.

Several recent studies have highlighted the advantages of WES over other techniques. Firstly, when applied to a heterogeneous cohort of undiagnosed HA patients, WES shows higher diagnosis efficiency (64%) (Pyle *et al.*, 2015) when compared to gene panels (< 20%) (Nemeth *et al.*, 2013). Different factors contribute to its efficiency. Firstly, HAs are characterized by an extensive clinical overlap between the different subtypes, especially within ADCAs. In this sense, gene panels are selected based on the observed clinical manifestations leading to misconceptions. By contrast, WES constitutes an unbiased sequencing technique detecting mutations found in coding regions, hence having a higher efficiency as it includes the known coding regions. In addition, it constitutes a cost-effective technique. Furthermore, the identification of a certain gene associated with distinct clinical manifestations extends its clinical phenotype. In other words, the genetic overlap between neurodegenerative diseases cannot be detected by using gene panels as they include genes strictly related to the target disease. However, these overlaps are perceived when using WES as it occurs in between HSP and HA, constituting a great breakthrough in research. Based on these observations HA and HSP are considered two different edges of a continuous spectrum disorder (Galatolo *et al.*, 2018) rather than two different pathologies.

To summarize, WES is a highly efficient technique that can be applied for molecular diagnosis in a heterogeneous cohort of undiagnosed patients suffering from HA and HSP. Furthermore, it permits to acquire a global insight regarding clinical and genetic overlapping between neurodegenerative disorders and therefore, these findings can be applied to other patients (Pyle *et al.*, 2015).

Finally, although nowadays there is not an effective treatment for neither HA nor HSP, ensuring a genetic diagnosis is highly relevant for genetic counseling and potential developing treatments.

2. Hypothesis and objectives

The determination of an accurate diagnosis in neurodegenerative diseases is challenging, as they constitute a miscellaneous group of disorders presenting overlapping clinical manifestations. HA and HSP are not the exception. Therefore, genetic analysis is required to confirm the diagnosis. In this sense, we hypothesize that WES analysis constitutes a powerful diagnostic tool when applied to patients presenting with adult-onset ataxia. WES allows the identification of mutations in the coding region of the genome and therefore, it is possible to detect the implication of genes previously associated with HA and/or HSP as well as the discovery of new genes.

The outcome obtained from WES analysis consists of a considerable catalogue of candidate mutations, which require further interpretation to determine their contribution to the clinical picture. In this sense, we hypothesize that the combined contribution of *in silico* analysis, segregation analysis and experimental approaches constitutes a promising tool to identify disease-causing variants.

Furthermore, unlike SCA1-3, SCA6, SCA7 and SCA17 genetic analysis, SCA36 expansion is not routinely performed in most genetic diagnostic laboratories. Consequently, undiagnosed SCA36 patients may exist.

Based on the aforementioned statements, the objectives of this Final Degree Project (FDP) are the following:

- **First objective:** Identification of the genetic cause of HA by WES data analysis from three undiagnosed families presenting with adult-onset ataxia.

- **Second objective**: Set up of a genetic analysis based on RP-PCR for efficient diagnosis of SCA36.

- **Third objective**: Determine the pathogenic contribution of *POLR3A* c.3688G>A variant, identified by gene panel, in the phenotype of a proband presenting with HSP.

3. Materials and methods

3.1 Whole exome sequencing

3.1.1 Family description

This FDP comprises three families, fATX-167, f-ATX-163 and fATX-166, with patients suffering a form of HA. For all the families, a relevant number of genes involved in ataxia were previously ruled out, by NGS (**Appendix I**).

The presence of pathogenic expansions in genes responsible for SCA1, SCA2, SCA3, SCA6, SCA7, SCA8, SCA12, SCA17, DRPLA and FXN were also discarded (**Appendix I**).

3.1.1.1 Family fATX-167

Patients of family fATX-167 (**Figure 3**) suffer from hereditary spinocerebellar ataxia and after analysis of candidate genes, remain without genetic diagnosis. In this study, WES data from three fATX-167's members were filtered and analyzed: the proband (SGT-1310), her healthy brother (SGT-1444), and an affected cousin (SGT-1445).

The proband (SGT-1310) is a 65-year-old woman diagnosed with spinocerebellar ataxia who presents with vertical nystagmus (involuntary and rapid eye movements) and axial ataxia (absence of static equilibrium). In addition, she has a sensitive neuropathy characterized by areflexia (absence of reflexes).

The proband has two affected sisters. In this sense, SGT-1493 is a 63-year-old woman who presents with a moderate apendicular ataxia, characterized by a decrease in coordination in the extremities, especially noticeable in arms and hands. In addition, she suffers from severe axial ataxia, severe dysarthria (difficulty in the speech) and both vertical and horizontal nystagmus.

SGT-1494 is a 60-year-old woman who presents with axial ataxia and therefore requires walking aid. Her first symptoms appeared at the age of 52 consisting of muscle stiffness and cramps; especially localized in feet, calves and hands. She also suffers from acute horizontal nystagmus and mild vertical nystagmus as well as dysarthria.

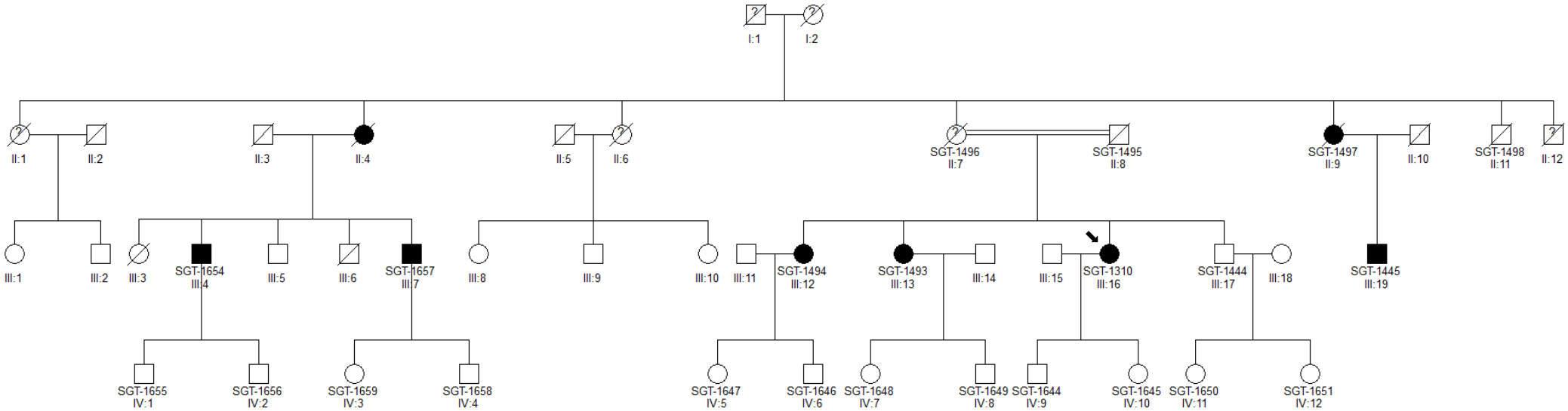

Figure 3: Family fATX-167's pedigree. Affected woman SGT-1310 is the proband (indicated with an arrow). Deceased individuals are indicated with a crossing line and the question mark indicates unclear clinical information.

3.1.1.2 Family fATX-163

Family fATX-163's patients (**Figure 4**) show a familiar cerebellar ataxia likely transmitted as an AD trait. WES data from three affected individuals were filtered and analyzed: SGT-1433 (proband), SGT-1459 and SGT-1432.

SGT-1433 (proband) is a 62 year-old-woman presenting with moderate axial ataxia and moderate tetrapendicular ataxia. The first clinical manifestations indicating cerebellar affection were detected at the age of 58. She also presents with nystagmus and mild to moderate dysarthria.

SGT-1432 is a 66-year-old man presenting with gait instability, thus requiring walking aids. The first clinical manifestations appeared at the age of 57. Furthermore, he suffers from incoordination in the limbs, especially affecting fine motor skills. He also presents with mild tetrapendicular dissymmetry and moderate axial ataxia although his reflexes remain active. Finally, mild dysarthria is also detected. His twin, SGT-1459 suffers from axial and tetrapendicular ataxia as well as moderate dysarthria.

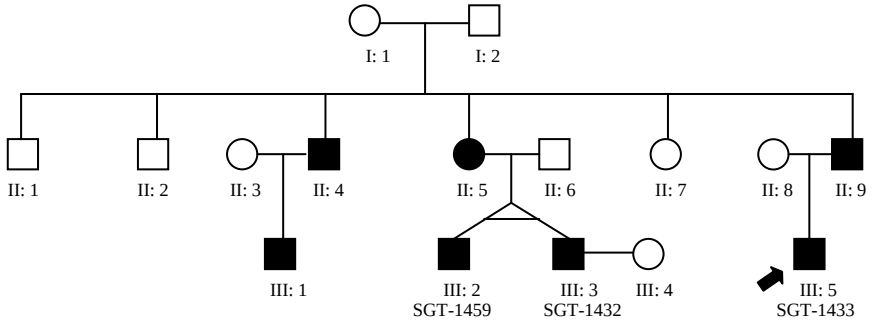

Figure 4: Family fATX-163's pedigree. Affected woman SGT-1433 is the proband (indicated with an arrow).

3.1.1.3 Family fATX-166

Family fATX-166's members have been diagnosed with late spinocerebellar ataxia with ocular apraxia showing an AR pattern of inheritance. In this case, WES was performed in the three individuals of the second generation: SGT-85 (proband), SGT-1457 (affected sister), and SGT-1458 (healthy brother) (**Figure 5**).

The proband (SGT-85) is a 56-year-old man presenting high expression levels of alpha-fetoprotein (AFP), thus indicating possible mutations in *SETX* (senataxine) gene that were previously discarded by Sanger sequencing. Mutations in *SETX* gene cause AOA2. In addition, he does not present dysarthria or dysphagia (difficulty in swallowing). He also suffers from transitory diplopia (double ocular perception).

SGT-1457 is a 47-year-old woman with severe gait instability and inability to walk, requiring a wheel chair since the age of 32. The first clinical manifestations appeared when she was 26. She also presents with mild dysarthria and mild to moderate apendicular dissymmetry. Although diagnosed with ocular apraxia she does not have retinitis pigmentosa, unlike her brother (SGT-1458). Furthermore, increased expression levels of in AFP were detected; nevertheless, the results came back negative as it occurred with SGT-85.

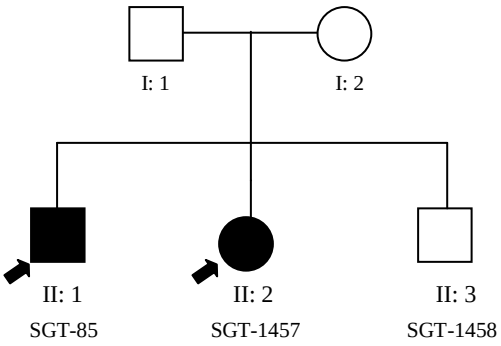

Figure 5: Family fATX-166's pedigree. Both patients have been clinically assessed and followed up for years. Hence; SGT-85 and SGT-1457 are considered probands in this family (indicated with an arrow).

3.1.2 WES and variant calling

WES of the different families as well as the variant calling process was performed at the Centre Nacional d'Anàlisi Genòmica (CNAG, Barcelona) which provided a VCF (Variant Calling Format) file containing the variants obtained. In addition, another annotated file was generated in the laboratory by using the Variant Effect Predictor (VEP) tool of Ensembl https://www.ensembl.org/info/docs/tools/vep/index.html). As a result of this process, a file containing information about the mutations was obtained, including: the chromosomal position, the transcript identification, the codon and amino acid change, as well as information from *in silico* predictors about the benignity of the mutation. In this procedure, the obtained sequences were aligned against the GRCh37/hg19 reference genome.

3.1.3 Filtering and candidate variant selection

After variant annotation in VEP file, mutations were filtered following a pipeline developed in the laboratory in order to narrow down the number of candidate variants associated with the disease.

Firstly, different files were generated in each of the families based on the type of expected inheritance observed in the pedigrees.

With respect to fATX-167 (**Figure 3**), several files were generated considering different patterns of inheritance. Firstly, an AD transmission was proposed even though patients suffering from ataxia are not present in all the different generations. Nevertheless, it must be taken into account that most individuals from I and II generations are deceased and detailed clinical information was not provided possibly because some of them shown minor symptoms. Secondly, an AR pattern of inheritance was considered. In addition, an X-linked mechanism of inheritance could not be dismissed because both males and females are affected (**Table 1**).

Table 1: Inheritance hypothesis for file generation of fATX-167

Sample	Family	Individual	AD	AR	X-linked
SGT-1310	fATX-167	proband (A)	0/1	1/1	0/1
SGT-1444		brother (H)	0/0	0/0-0/1	0/0
SGT-1445		cousin (A)	0/1	-	1/1

0/0: absence of the mutation. 0/1: mutation in heterozygosis. 1/1: mutation in homozygosis.
A: affected individual. H: healthy individual.

Regarding fATX-163 (**Figure 4**), an AD pattern of inheritance is expected due to the fact that affected individuals are present in the three generations, and hence, an AR transmission was ruled out. X-linked pattern of inheritance can be dismissed by observing the inheritance between individual II: 4 and III: 1, which are both males. Therefore, in fATX-163 a single file considering AD inheritance was generated (**Table 2**).

Table 2: Inheritance hypothesis for file generation of fATX-163.

Sample	Family	Individual	AD
SGT-1433	fATX-163	proband (A)	0/1
SGT-1432		cousin (A)	0/1
SGT-1459		cousin (A)	0/1

0/1: mutation in heterozygosis. A: affected individual.

Finally, with regard to fATX-166 (**Figure 5**) a single file was generated considering an AR fashion of inheritance due to the fact that neither of the progenitors is affected but two out of three individuals in the second generation are. Following this procedure, because of the absence of a homozygous candidate mutation, the compound heterozygosis hypothesis was considered (**Table 3**). According to this hypothesis, affected individuals (SGT-85 and SGT-1457) carry two heterozygous mutations in the same gene. Therefore, SGT-1458 (non-affected) could carry a single mutation (0/1) or neither one (0/0).

Table 3: Inheritance hypothesis for file generation of fATX-166.

Sample	Family	Individual	AR	Comp. Het.
SGT-85	fATX-166	proband (A)	1/1	0/1
SGT-1457		sister (A)	1/1	0/1
SGT-1458		brother (H)	0/0-0/1	0/0-0/1

0/0: absence of mutation. 0/1: presents the mutation in heterozygosis. 1/1: mutation in homozygosis. A: affected individual. H: healthy individual. Comp. Het.: Compound Heterozygous.

Then, the candidate mutations were selected following different criteria. In this sense, the chosen mutations include those that encode proteins or that alter the splicing.

In addition, variants presenting a general MAF (Minor Allele Frequency) >0.005 according to The Genome Aggregation Database (gnomAD: https://gnomad.broadinstitute.org/) were discarded. GnomAD is a database that contains exome and genome sequencing data of thousands of unrelated healthy individuals.

Finally, variants were selected according to the number of alleles found in control population according to the criteria explained using the Sherloc (Semiquantitative, Hierarchical Evidence-based Rules for Locus Interpretation) system (Nykamp *et al*., 2017).

These criteria are based on the quality and abundance of the Exome Aggregation Consortium (ExAC) data at the locus, as well as the expected pattern of inheritance of the gene, and the frequency of the variant in ExAC. Following Sherloc, when a mutation is found in a locus presenting high quality and abundant data (> 15000 alleles in ExAC), the mutation is probably pathogenic if the frequency of the variant in ExAC is ≤8 alleles. If the variant is absent in ExAC the mutation is likely damaging. Consequently, novel variants and those presenting less than 10 alleles in control population were selected as possibly deleterious changes and followed the prioritization criteria. Furthermore, when having AR inheritance, the presence of homozygote individuals in control population dismisses the harmful character of the mutation, and hence, those variants are not further considered in the study.

3.1.4 Variant prioritization

After variants have been selected according to the number of alleles found in control population other parameters have to be taken into account for prioritization, including: the gene containing the mutation, the sequencing coverage, whether the mutation is associated with a disease, the protein's function, the level of protein expression in clinically relevant organs and the consequence of the mutation. In this sense, different databases are used in order to acquire a deep insight of each of the selected mutations.

Firstly, both the coverage and the region containing the mutation are analyzed to identify possible artifacts during sequencing. In this sense, if more than 5 mutations are identified within a region of less than 20 bp; some of them are presumed artifacts due to the characteristics of the region.
Secondly, the consequence of the mutation is of great relevance, prioritizing frame-shift or missense mutations over synonymous ones. In addition, *in silico* predictors are used in order to predict the pathogenic effect of missense mutations based on multiple sequence alignments.

Polyphen-2 (http://genetics.bwh.harvard.edu/pph2/) is based on position-specific independent counts (PSIC) score when the mutation is in a functional domain. This score takes into account the statistical significance of changes between aligned sequences and the relative position of the mutation within the sequence. By contrast SIFT (Sorting Intolerant From Tolerant; https://sift.bii.a-star.edu.sg/) is based on sequence-homology between amino acids necessary for the protein's function, assuming their conservation through evolution. MutationAssesor.org (http://mutationassessor.org/r3/) was also used in order to determine the functional impact of missense mutations in proteins, based on the degree of amino acid conversation in protein homologs.

Regarding splicing changes and synonymous variants, three *in silico* predictors were used to analyze the possible alteration of splicing: HSF (Human Splicing Finder; http://www.umd.be/HSF/), NetGene (http://www.cbs.dtu.dk/services/NetGene2/) and NNSplice (https://www.fruitfly.org/seq_tools/splice.html).

Furthermore, OMIM (Online Mendelian Inheritance in Men; https://www.omim.org/) and HGMD Professional (Human Genetics Mutation Database; http://www.hgmd.cf.ac.uk/ac/index.php) were used to determine whether the candidate mutations/genes have already been associated with a disease and its type of inheritance. As a result, variants associated with neurodegenerative diseases/phenotypes and with the type of inheritance expected from each of the families are of great interest.

Information regarding protein's function was obtained from UniProt (https://www.uniprot.org/). On the other hand, The Human Protein Atlas (HPA; https://www.proteinatlas.org/) enables to determine the degree of expression of mRNA and proteins in different tissues and organs.

In addition, MGI database (Mouse Genome Informatics; http://www.informatics.jax.org/) was consulted in order to obtain information concerning the gene of interest in mouse models. This database provides information not only about the level of protein's expression in different tissues but also about the implications of the gene in different disease models.

3.1.5 Validation and segregation by Sanger sequencing

In order to validate the selected mutations in each family, Sanger sequencing was performed following PCR reaction.

3.1.5.1 Primer design

In order to amplify the region containing the variation primers are designed following the genome sequences found in Ensembl Genome Browser 37 (Ensembl: http://grch37.ensembl.org/index.html).

Primers were designed for validation of each of the selected mutations following several criteria shown below. In addition, their physicochemical properties were evaluated by using the tool OligoEvaluator of Sigma-Aldrich (http://www.oligoevaluator.com/LoginServlet).

1. Primers' length should be in a range of 18-22 bp in order to provide enough specificity during amplification at the same time it enables easy binding to the template region. In addition, minimum length difference between them must be ensured.

2. Primers' temperature of melting should be around 60ºC. The difference between the primers' temperature should not be higher than 2ºC.
3. The amplicon length should range between 300-700 bp.
4. Autocomplementary sequences should be avoided in order to prevent the formation of secondary structures.
5. Frequent polymorphisms, especially SNPs cannot be included in the primers' hybridization region; otherwise, there is a risk of losing an allele during amplification.
6. In order to improve hybridization, primers present in the 3' end a high content in CG.
7. To perform Sanger sequencing it is preferable to design primers more than 50 bp away from the mutation as the first nucleotides read are of lower quality.

Finally, once primers have been designed, their specificity has to be proved through the Primer-BLAST tool (https://www.ncbi.nlm.nih.gov/tools/primer-blast/).

3.1.5.2 PCR and Sanger sequencing conditions

PCR reactions were performed by using the Veriti® Thermal Cycler (Applied Biosystems, Waltham, MA, USA). The primers used for each of the families, together with their hybridization temperature and the amplicon length, are shown in **Appendix II.** Furthermore, the PCR reagents and conditions are indicated in **Appendix II**.

PCR products were tested by electrophoresis in agarose gel at 1% and using GelRed (Biotium Fremont, MA, USA) as an intercalating agent. Agarose gels were prepared by using TAE 1% (Tris, acetate, EDTA, H_20) and low-EEO agarose (Scharlab SL, Barcelona, Spain). The molecular weight marker used was 1Kb Plus DNA Ladder (Thermo Fisher Scientific, Waltham, MA, USA).

Sanger sequencing was performed by the Service of Translational Genetics and Genomics (STGG) from CIPF using a 3730xl DNA Analyzer (Applied Byosistems, Waltham, MA, USA). The obtained electropherograms were analyzed with the software Chromas (version 2.5). Finally, the sequences were compared with the reference genome using BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

3.2 SCA36 screening of a patient cohort

3.2.1 Patient cohort description

A cohort of 52 patients was studied, including patients from families fATX-163, fATX-166 and fATX-167. **Table 4** shows the individuals investigated from each family. All patients included in the SCA36 screening presented with ataxia of adult-onset and were negative for the previous genetic tests indicated in **Appendix I.**

Table 4: Individuals from families fATX-163, fATX-166, and fATX-167 included in the SCA36 screening.

Patient ID SGT-	1432	1433	1459	85	1457	1493	1494
Family	fATX-163			fATX-166		fATX-167	

3.2.2 Set up of SCA36 screening with positive controls

In order to perform the SCA36 screening, two different PCR reactions were performed: a standard PCR and a repeat-primed PCR (RP-PCR). The first one enables the amplification of non-expanded alleles whilst the second allows the amplification of the alleles containing the hexanucleotide expansion GGCCTG, found in *NOP56* gene.

The conditions for both reactions were determined by using three positive SCA36 patients as positive controls.

3.2.3 Primer design

All primers used in this study, with the exception of SCA36_R, were previously described (Kobayashi *et al.*, 2015). Primer SCA36_R was designed following the same criteria described in Section 3.5.1.1.

The amplicon length of the products obtained by standard PCR can be determined by testing and fragment analysis. By contrast, neither the number of repeats nor the length of the expanded alleles can be established. The expected amplicon length, obtained from the reference genome, is 174 bp (Ensembl Genome Browser 37).

Three primers have been used in the RP-PCR (**Table 5**):

- SCA36_F: complementary to a region near the SCA36 repeat that hybridizes with the template DNA. This primer is labeled with a fluorescent flag, 6-FAM (6-carboxyfluorescein), in the 5' end in order to detect the PCR products during fragment analysis.
- SCA36_Anchor: complementary to a region of SCA36_RV1 primer.
- SCA36_RV1: containing two different regions. Firstly, towards the 5' end, it presents a region (highlighted in **table 5**) that does not hybridize with the template but it is complementary to SCA36_Anchor. Secondly, a region complementary to the repeat that will hybridize with the template.

Therefore, the SCA36_RV1 primer can bind to different regions of the hexanucleotide repeats leading to the obtention of fragments of variable lengths depending on the binding site. Furthermore, SCA36_Anchor primer will lead to further amplification of these fragments in order to obtain a stronger signal that will appear as a stairway.

Table 5: Primers for SCA36 screening.

Gene	PCR type	Primers	Sequencies 5'-3'	Hybridization Tº (ºC)
NOP56	**Standard PCR:** Alleles amplification	SCA36_F	TTTCGGCCTGCGTTCGGG	65
		SCA36_R	TAGCCGACCGCGTGCTCAA	
	Repeat-Primed PCR: Expansion amplification	SCA36_F	TTTCGGCCTGCGTTCGGG	56
		SCA36_Anchor	TACGCATCCCAGTTTGAGAC	
		SCA36_RV1	**TACGCATCCCAGTTTGAGAC**GCAGGCCCAGGCCCAGGCCCAGGCC	

3.2.4 PCR reactions

Two PCRs were carried out in all patients. Firstly, standard PCR followed by RP-PCR. Both were performed in a ProFlex PCR System thermocycler (Applied Biosystems, Waltham, MA, USA).

3.2.4.1 Standard PCR: Alleles amplification

Standard PCR conditions and reagents are presented **Appendix III**. The expansion step is 30 seconds long, and therefore, non-expanded alleles can be amplified. Nevertheless, the products of this reaction vary in length depending on the number of repeats that each patient presents. The products obtained from this reaction were tested using the automated electrophoresis tool QIAxcel Advanced (QiaGen, Hilden, Germany).

3.2.4.2 Repeat Primed-PCR: Expansion amplification

RP-PCR conditions are shown in **Appendix III.** Several parameters have been modified in order to ensure an efficient amplification of the region, due to the number of repeats, the length of the expected products and the rich GC content. The hybridization temperature has been reduced up to 56ºC and the extension time is longer (2 min). In addition, the number of

cycles of the second step has been increased to 45 because two consecutive amplifications take place. Firstly, the amplification of the template using SCA36_RV1 primer from which products of variable lengths is obtained. Secondly, further amplification of the products obtained in the former, through SCA36-anchor primer. The latter, enables to increase the strength of the signal for further fragments analysis.

Reagents used in RP-PCR are indicated in **Appendix III**. DMSO was found to be crucial to ensure product amplification. Furthermore, a higher DNA concentration is also required in comparison with the standard PCR. The products obtained from this reaction were directly analyzed by capillary electrophoresis using a HITACHI 3130xl Genetic Analyzer (Applied Biosystems, Waltham, MA, USA).

3.2.4.3 Fragment analysis

Fragment analysis is a technique based on capillary electrophoresis, which separates fluorescently labelled PCR products based on its length with high accuracy.

The products obtained from the standard PCR were diluted to different ratios depending on the strength of the signal obtained from the initial testing (Section 3.2.4.1). Following this procedure, the samples obtained from both reactions were processed in the same way.

1 µL of PCR product (or its dilution), 12 µL of formamide and 0.5 µL of GeneScan™500ROX™ (Applied Biosystems, Waltham, MA, USA) were mixed. Formamide is used in order to ensure the migration of the DNA as a single chain which avoids the formation of secondary structures. On the other hand, ROX500 is used as a molecular marker. Finally, the results were analyzed with the program GeneMapper (Applied Biosystem, Waltham, MA, USA).

3.2.4.4 Determination of repeat number

The number of repeats in non-expanded alleles can be calculated; having its length and the amplicon length (minus the number of bp constituting the repeats).

$$number\ of\ repeats = \frac{allele\ length - (amplicon\ length - bp\ of\ the\ repeat)}{repeat\ length}$$

Where:

- Allele length: data obtained from fragment analysis.
- Amplicon length: 174 bp (Section 3.2.3).
- Base pairs of the repeat: number of bp within the amplicon length corresponding to the hexanucleotide repeat = 42 bp. The reference genome (Ensembl Genome Browser 37) has 7 repeats.
- Repeat length = 6 bp (GGCCTG). In some patients variability is observed, detecting CGCCTG expansion.

3.3 Analysis of the *POLR3A* c.3688G>A variant

POLR3A encodes for the largest catalytic subunit of RNA polymerase III. Recently identified mutations have broadened the phenotype associated to this gene; specifically, c.1909+22G>A mutation (intron 14) has been related to HSP and CA due to the activation of a cryptic splice site (Minnerop *et al.,* 2017). Commonly, affected individuals were compound heterozygotes presenting this mutation in at least one allele and a wide range of variable mutations in the other one.

Based on these observations, the objective relied on the establishment of a genetic diagnosis for SGT-1426 (proband). He is a 33-year-old man presenting with slowly progressive and sporadic HSP whose first symptoms appeared at the age of 22. He suffers from gait and balance

instability as well as slight dysarthria. Furthermore, sensitivity loss was detected due to alterations in the dorsal root nerves which constitute the afferent sensory root of spinal nerves. He also presents with affection of spinocerebellar tracts and mild cerebellar ataxia. Finally, MRI results revealed hyperintensities in the midbrain.

The patient was screened for candidate mutations by using AtxSPG-365 gene panel designed by the Unit of Genetic and Genomics of Neuromuscular and Neurodegenerative Disorders. This custom-made panel is made using SureSelectXT technology (Agilent Technologies In., SantaClara, CA, USA) targeting HiSeq (Illumina In, San Diego, CA, USA) and contains 365 genes associated with HA and HSP (**Appendix IV**).

Sequencing data obtained from AtxSPG-365 panel was analyzed and putative mutations in *POLR3A* gene were detected. In this sense, SGT-1522 (unaffected father) and SGT-1426 (proband) were compound heterozygous for the c.1909+22G>A intronic variant and the c.3688G>A (p.D1230N) missense mutation (**Figure 6**). Furthermore, *in silico* analysis was performed to identify whether this may alter the proper splicing (**Appendix V**). Finally, mRNA from peripheral blood was extracted from SGT-1426 to perform a transcript analysis, and further, confirm the *in silico* analysis.

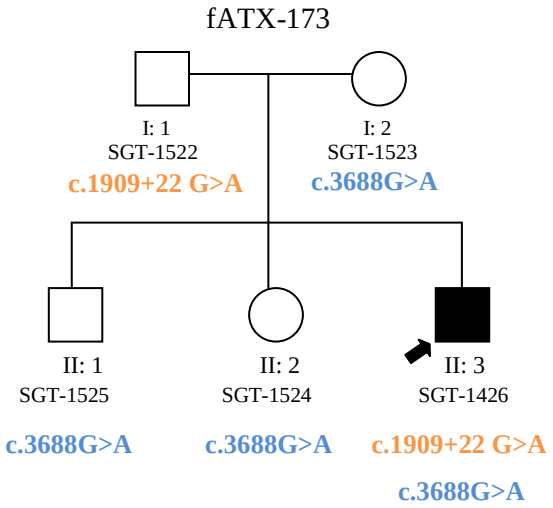

Figure 6: fATX-173's pedigree. Affected man SGT-1426 is the proband (indicated with an arrow).

3.3.1 mRNA extraction and retrotranscription to cDNA

In order to perform a functional analysis, mRNA was extracted from a sample of peripheral blood from two individuals: a control and the affected subject SGT-1426. The commercial kit PAXgene Blood RNA Kit IVD (QiaGen, Hilden, Germany) was used.

Once mRNA was extracted, retrotranscription was performed by using the commercial kit: cDNA SuperMix (Quantabio, Beverly, MA, USA). The reaction was performed in a 2720 Thermal Cycler (Applied Biosystems, Waltham, MA, USA) and the RT-qSCRIPT Programme was used: (1) 5 min at 25ºC; (2) 30 min at 42ºC; (3) 5 min at 85ºC; (4) Storage at 4ºC.The reagents used are indicated in **table 6**.

Table 6: Reagents and volumes used for retrotranscription reaction.

Volume (µl)	Control	SGT-1426
qScript cDNA SuperMix	4	4
RNA	2.5	5
Free RNAase and DNAase H_2O	13.5	11
Final volume	20	20

In order to perform efficiently the retrotranscription reaction, 500 ng of RNA are needed. Therefore, the RNA volume introduced depends on the concentration of mRNA that was extracted, which was calculated in a NanoDrop system (Thermo Fisher Scientific, Waltham, MA, USA).

3.3.2 Quality control: GAPDH amplification

Following retrotranscription, a quality control is performed in order to determine whether the reaction was done successfully and to identify quantitative differences in the relative expression between the control and SGT-1426. In this sense, a PCR targeting the amplification of *GAPDH* gene is done. *GAPDH* presents a ubiquitous expression and therefore it is present in the peripheral blood sample and can be used as a quality control. PCR conditions and the program are described in **Appendix II**.

In addition, the primers and the hybridization temperature are presented in **table 7**, and the PCR products were tested by electrophoresis following the conditions presented in 3.1.5.2.

Table 7: Primers used in *GAPDH* cDNA amplification.

Gene	Primers	Sequencies 5'-3'	Amplicon length (bp)	Hybridization Tº (ºC)
GAPDH	GAPDH_F	CATTGACCTCAACTACATGG	395	58
	GAPDH_R	CAAAGTTGTCATGGATGACC		

3.3.3 Primer design, PCR reaction and Sanger sequencing

The objective of this procedure relies on the identification of alterations in the splicing caused by *POLR3A* c.3688G>A mutation, and therefore, cDNA was used as template. The transcript *POLR3A*-002 (ENSP00000361446.3, NM_007055.4) was used as reference and its sequence was obtained from Ensembl Genome Browser 37 GRCh37 (http://grch37.ensembl.org/Homo_sapiens/Gene/SuMary?db=core;g=ENSG00000148606;r=10:79734907-79789303).

The primers were designed following the same procedure presented in Section 3.1.5.1. Moreover, based on bibliographic research, exon skipping and significant deletions were common mechanisms present in *POLR3A* when it comes to alteration of splicing. Consequently, as the target mutation is located in exon 28, primers were designed to hybridize in exons 26 and 30 in order to identify possible deletions in exons 27 and 29. The primers used and their annealing temperature are presented in the following table (**Table 8**).

Table 8: Primers used in *POLR3A* cDNA amplification.

Gene	Primers	Sequencies 5'-3'	Amplicon length (bp)	Hybridization Tº (ºC)
POLR3A	POLR3AcDNA_E26_F	CCTGGAACGGATTAGGCTTCT	522	60
	POLR3AcDNA_E30_R	CCAAACCTAGTGATGCCCA		

Additionally, cDNA from both SGT-1426 and the control were amplified in order to compare the pattern of bands observed in electrophoresis. PCR conditions were the same as in **table 6** but supplemented with 0.5 µL of Mg^{+2}. PCR products were tested following the same procedure as in Section 3.1.5.2.

Sequencing conditions were the same as presented in Section 3.1.5.2. Furthermore, PCR products were sequenced using both forward and reverse primers.
Finally, the pattern of alteration of the splicing was analyzed manually by observing the electropherogram provided in the sequencing reaction.

3.3.4 cDNA purification from agarose gel

The difference in length found between the wild-type and the mutant transcript was of 74 bp. Consequently, in order to facilitate DNA purification of both bands three PCR products were loaded together in a single bigger well in a double agarose gel prepared at 2.5%. The agarose was prepared with TAE 1% (Tris, acetate, EDTA, H_20). Samples were loaded using the 6x Purple Gel Loading Dye buffer (New England Biolabs, Ipswich, MA, USA).The samples were running at 100V for 2 hours. After, the gel was introduced in a solution of GelRed (Biotium, Fremont, MA, USA) and maintained in agitation for 30 min in order to perform the staining of the samples.

Following this procedure, the two bands were cut by using a UV transilluminator and DNA was extracted by using the extraction kit NucleoSpin® Gel and PCR Clean-up (MACHEREY-NAGEL GmbH & Co. KG, Düren, Germany). Finally, the bands were sequenced as described in Section 3.1.5.

4. Results

4.1 Whole exome sequencing

The WES data obtained from three families (fATX-167, fATX-163 and fATX-166) was filtered following the inheritance hypothesis described in Section 3.1.3. Then, several candidate mutations were identified in each family as they accomplished the variant prioritization criteria presented in Section 3.1.4.

4.1.1 Family fATX-167

In family fATX-167 (**Figure 3**), as indicated in Material and Methods, exome of three subjects were sequenced: the proband (SGT-1310), her healthy brother (SGT-1444), and an affected cousin (SGT-1445).

Four novel synonymous variants, in principle inherited in an AD manner, were detected by WES: (1) *GPR111* c.666A>G (p.S222S); (2) *DMBT1* c.246T>C (p.I82I); (3) *RPGRIP1L* c.1434C>G (p.S478S); and (4) *FAM222B* c.219C>T (p.H73H). The *in silico* tools predicted that none of these variants may alter the proper splicing, and accordingly, all of them were ruled out as disease-causing mutations.

Four promising novel variants were identified by WES. The sequencing results as well as the phenotypes of each of the individuals belonging to fATX-167 are presented in **table 9**.

Table 9: Novel candidate missense mutations for fATX-167 from WES analysis.

Family	Gene	Individual	Result	cDNA change /Protein change	*In silico* predictors	
					SIFT	Polyphen-2
fATX-167	*SLC38A7* (NM_018231.3)	SGT-1310	Heterozygous	c.*37G>A/-	T	B
		SGT-1444	Negative			
		SGT-1445	Heterozygous			
	CFL2 (NM_138638.5)	SGT-1310	Heterozygous	c.276A>C/p.K92N	T	PD
		SGT-1444	Negative			
		SGT-1445	Heterozygous			
	BOD1L1 (NM_148894.3)	SGT-1310	Heterozygous	c.5302G>C/p.P1768A	T	B
		SGT-1444	-			
		SGT-1445	Heterozygous			
	MUTYH (NM_001128425.1)	SGT-1310	Heterozygous	c.1169C>T/p.V390A	D	PD
		SGT-1444	Negative			
		SGT-1445	Heterozygous			

T: tolerated, B: benign, D: deleterious, PD: probably damaging. NM_X: Indicates the reference transcript variant used for each of the genes.

***SLC38A7* c.*37G>A.** *SLC38A7* encodes the sodium-dependent amino acid transporter (family 38 member 7) which binds preferentially to L-glutamine. The protein is highly expressed in the brain although it is found ubiquitously (HPA). Within the brain, it is detected specifically in neurons (Hägglund *et al.*, 2011). It is a novel missense mutation leading to the substitution of glycine (neutral) by glutamic acid (negatively charged). The *in silico* predictors classified the mutation as tolerated (SIFT) and benign (Polyphen-2). So far, no disease is related to mutations in *SLC38A7*. The segregation analysis revealed that the patient SGT-1494 does not harbor the aforementioned mutation and therefore, it can be discarded (**Table 10**).

***CFL2* c.276A>C (p.K92N).** *CFL2* encodes the cofilin-2 protein involved in the actin polymerization and depolymerization in muscles where it presents its highest expression (HPA). It is a novel missense mutation leading to the substitution of lysine (positively charged) by asparagine (neutral). In this sense, it was classified as tolerated (SIFT) and as probably damaging (Polyphen-2) by the *in silico* predictors. Regarding Mutation Assessor predictor, this

variant has a low impact on the protein functionality. Mutations in *CFL2* were associated with nemaline myopathy (MIM 610687), which is characterized by muscle weakness, hypotonia and delayed motor development. Importantly, the *CFL2*-associated phenotype does not fit with the fATX-167's clinical picture. Finally, the co-segregation study identified two affected individuals (SGT-1493 and SGT-1494) that did not carry the mutation. Consequently, this candidate mutation was dismissed (**Table 10**).

***BOD1L1* c.5302C>G (p.P1768A).** *BOD1L1* codifies for the orientation of chromosomes in cell division protein, therefore it is expressed ubiquitously (HPA). It is a novel missense mutation that causes the substitution of proline by alanine. Regarding the *in silico* predictors, it is described as tolerated (SIFT) and benign (Polyphen-2) and it was predicted to have a medium impact on the protein functionality (Mutation Assessor). To date, no disorder is associated with *BOD1L1* mutations. Moreover, the co-segregation study was performed exclusively in the affected individuals of the family because the mutation was considered as a possible artifact, since a different mutation, closely localized, was found in an unrelated patient who presents with a disparate movement disorder (NBIA, Neurodegeneration with Brain Iron Accumulation). SGT-1493 (affected) did not present the mutation and consequently, it was ruled out as a candidate causative variant (**Table 10**).

Table 10: Segregation analysis in fATX-167 for mutations in *SLC38A7, CFL2* and *BOD1L1*.

Individual	Status	*SLC38A7*	*CFL2*	*BOD1L1*
SGT-1310	Affected	P	P	P
SGT-1444	Healthy	N	N	-
SGT-1445	Affected	P	P	P
SGT-1493	Affected	P	N	N
SGT-1494	Affected	N	N	P
SGT-1495	Healthy	N	N	-
SGT-1496	Healthy	P	P	-
SGT-1497	Affected	-	-	P
SGT-1498	Healthy	-	N	-

P: positive; N: negative. In red, non-concordant findings.

***MUTYH* c.1169T>C (p.V390A)**. *MUTYH* encodes a DNA glycosylase, involved in DNA repair generated under oxidative stress conditions. In the presence of reactive oxygen species (ROS), 8-oxoguanine (8-oxoG) is obtained and this modified guanine misspairs with adenine, leading to the substitution of C-G by T-A. This protein is expressed both in the nucleus and in the mitochondria.

Regarding the mutation, it is a novel missense one, substituting valine by alanine and it has been classified as deleterious (SIFT) and as probably damaging (Polyphen-2) by the *in silico* predictors. In Mutation Assessor database it was determined that the mutation would lead to a medium impact on the functionality of the protein. *MUTYH* c.1169T>C is a novel mutation; nevertheless in HGMD Professional, there are many pathogenic mutations described in the region near the candidate one. The majority of them in OMIM and HGMD Professional have been associated with colorectal cancer or *MUTYH*-associated polyposis. However, several studies have correlated the presence of 8-oxoG and *MUTYH* base excision DNA repair (BER) mechanism with neurodegeneration (Sheng *et al.*, 2012). Neurons are affected by 8-oxoG accumulation in mitochondrial DNA (mtDNA), whereas microglia is affected by accumulation in nuclear DNA (nDNA). Although the mechanism leading to neurodegeneration remains uncertain, it has been observed that suppression of *MUTYH* expression leads to protection from neurodegeneration under oxidative stress conditions (Sheng *et al.*, 2012). This protein is expressed ubiquitously, nevertheless within the brain; it presents the highest expression in the cerebellum (HPA).

In this case, the segregation analysis revealed some controversial findings (**Table 11**):

- SGT-1496 carries the *MUTYH* c.1169T>C change. Unfortunately, this woman deceased and no clinical information is available. Likely, this woman was asymptomatic or with very mild symptoms. We cannot forget that this form of ataxia is characterized by a late-onset in the fifties-sixties.

- SGT-1457 does not harbor the *MUTYH* c.1169T>C change. SGT-1457 is an affected man with a well-established clinical phenotype. Accordingly, *MUTYH* does not seem to be the gene responsible for ataxia in fATX-167.

Table 11: Segregation analysis in fATX-167 for c.1169T>C mutation in *MUTYH*.

Individual	Status	Segregation
SGT-1647	Healthy	P
SGT-1648	Healthy	N
SGT-1649	Healthy	N
SGT-1650	Healthy	N
SGT-1651	Healthy	P
SGT-1654	Affected	P
SGT-1655	Healthy	P
SGT-1656	Healthy	P
SGT-1657	Affected	N
SGT-1658	Healthy	N
SGT-1659	Healthy	N

Individual	Status	Segregation
SGT-1310	Affected	P
SGT-1444	Healthy	N
SGT-1445	Affected	P
SGT-1493	Affected	P
SGT-1494	Affected	P
SGT-1495	Healthy	N
SGT-1496	Healthy (?)	P
SGT-1497	Affected	P
SGT-1498	Healthy	N
SGT-1644	Healthy	N
SGT-1645	Healthy	N
SGT-1646	Healthy	N

P: positive; N: negative. In red, non-concordant findings. The question mark indicates unclear clinical information.

4.1.2 Family fATX-163

In family fATX-163 (**Figure 4**), as we indicated in Material and Methods, the three available samples (SGT-1459, SGT-1432 and SGT-1433), belonging all of them to affected individuals, were investigated by WES. No additional DNAs were available, so further segregation analysis could not be carried out. Moreover, because disease seems to be transmitted in an AD fashion, rare/novel heterozygous changes were prioritized. Novel variants were not identified. Regarding rare mutations, two changes deserve to be mentioned (**Table 12**):

Table 12: Candidate heterozygous mutations for fATX-163 from WES analysis.

Family	Gene	Individual	Result	cDNA change /Protein change	Frequency	*In silico* predictors	
						SIFT	Polyphen-2
fATX-163	*HPCAL1* (ENST00000422133)	SGT-1432	Heterozygous	c.275G>A/p.G92D	1 allele	D	B
		SGT-1433	Heterozygous				
		SGT-1459	Heterozygous				
	EPPK1 (NM_031308.4)	SGT-1432	Negative	c.6619A>C/p.T2207P	rs782630329	D	B
		SGT-1433	Negative				
		SGT-1459	Negative				

T: tolerated, B: benign, D: deleterious, PD: probably damaging. NM_X: Indicates the reference transcript variant used for each of the genes. RS represents the mutation identifier in Ensembl Genome Browser 37.

***HPCAL1* c.275G>A (p.G92D).** *HPCAL1* codifies for hippocalcin-like protein 1, a neuron specific Ca^{+2} binding protein. According to the HPA, the highest expression is found in the brain, especially in the cerebellum, concretely in the PC. The mutation is annotated in the gnomAD database with a very low frequency: MAF= $6.76x10^{-6}$ (1 allele / 148010 alleles). *HPCAL1* mutations have not been associated with a disease in neither HGMD nor OMIM databases. The mutation p.G92D is missense, leading to the substitution of glycine (neutral) by aspartate (negatively charged), predicted as deleterious by SIFT and benign by Polyphen-2. The three patients were carriers of the *HPCAL1* c.275G>A in heterozygosis, and hence, on the available information, this rare change could be involved in the clinical outcome in the family fATX-163.

***EPPK1* c.6619A>C (p.T2207P)**. *EPPK1* encodes the epiplakin protein, a cytoskeletal linker protein that connects to intermediate filaments and which is barely expressed in the brain (HPA). This mutation was classified as deleterious (SIFT) and benign (Polyphen-2). In addition, in Mutation Assessor, the variant was determined to have no impact on the protein functionality. Missense mutations in this gene are described in HGMD Professional contributing to susceptibility towards developmental disorders and early onset PD. Finally, this mutation has already been described (rs782630329), although linked to poor quality parameters (gnomAD). Moreover, the *EPPK1* c.6619A>C mutation was found in an unrelated family with NBIA, suggesting this variant may not be the causative mutation in fATX-163. The analysis of *EPPK1* c.6619A>C by Sanger sequencing showed that none of the three investigated patients really harbored the mentioned change and eventually, *EPPK1* c.6619A>C was ruled out as possible mutation causing disease in fATX-163.

4.1.3 Family fATX-166

Regarding family fATX-166 (**Figure 5**), as described in Material and Methods, samples from the three individuals of the II generation were studied by WES: SGT-85 (proband), SGT-1457 (sister) and SGT-1458 (brother). As indicated in fATX-163, no additional DNAs were available, so segregation analysis could not be performed. Furthermore, based on the family's pedigree (**Figure 5**) an AR pattern of inheritance is expected so rare homozygous mutations were prioritized. Nevertheless, the entire set of candidate variants obtained presented a MAF>0.005 (gnomAD) and therefore they were dismissed.

Then, a second annotation file was generated considering compound heterozygosis hypothesis (**Table 3**). The objective was to identify two heterozygous candidate mutations in the same gene, in this sense a candidate gene deserves to be mentioned:

***SHQ1* c.1031A>G and c.66C>T:** *SHQ1* encodes for a protein involved in the accumulation of H/ACA ribonucleoproteins, required for ribosomal synthesis through pseudophosphorylation of rRNA. This protein presents a ubiquitous expression; presenting the highest expression in pancreas, gastrointestinal tract and the testis. Within the brain it has been detected exclusively in the cerebral cortex (HPA). Mutations in this gene have not been associated with a disease in neither HGMD Professional nor OMIM. Two different heterozygous mutations were identified (**Table 13**):

- c.1031A>G (p.K338R): It is a missense mutation that has already been annotated in Ensembl Genome Browser 37 (rs749525383). In gnomAD database presents a MAF= $1{,}99x10^{-5}$ (5 alleles/ 251436 alleles). The mutation has been described as tolerated (SIFT) and probably damaging (Polyphen-2). Regarding Mutation Assessor database it was determined to have a medium impact on the protein functionality.
- c.66C>T (p.P22P): This is a synonymous mutation already annotated in Ensembl Genome Browser 37 (rs974621795). It has been classified as tolerated (SIFT) and probably damaging (Polyphen-2). Then, the *in silico* predictors described in section 3.1.4 were used to identify possible alterations of the splicing caused by this mutation. HSF software determined a putative anomalous splicing produced by an alteration of an exonic splice enhancer site. By contrast, both NetGene2 and NNSplice found no alterations of the splicing between the wild type and c.66C>T alleles (**Appendix VI**).

Table 13: Candidate compound heterozygous mutations for fATX-166 from WES analysis.

Family	Gene	cDNA change /Protein change	Frequency	Type of change	*In silico* predictors	
					SIFT	Polyphen-2
fATX-166	*SHQ1* (NM_018130)	c.1031A>G/p.K338R	5 (rs749525383)	missense	T	PD
		c.66C>T/p.P22P	0 (rs974621795)	synonymous	T	PD

T: tolerated, PD: probably damaging. NM_X: Indicates the reference transcript variant used for each of the genes. RS represents the mutation identifier in Ensembl Genome Browser 37.

This variant was not validated by Sanger sequencing, and therefore, it cannot be confirmed whether the affected individuals present the mutation.

Two or more heterozygous mutations were found in other seven genes (**Appendix VII**). Nevertheless they were ruled out for disparate reasons:

- *CELA3A:* This gene encodes a protease, which is exclusively expressed in the pancreas (HPA). So far, variants in this gene have not been associated with any disease in OMIM or HGMD Professional. In addition, both of the detected mutations present more than eight alleles in control population (gnomAD), and according to Sherloc criteria (Nykamp *et al.*, 2017) (Section 3.1.3) they are probably benign. Consequently, this gene is probably not involved in the family's phenotype.
- *ALG6:* encodes the alpha-1,3-glucosyltransferase which is ubiquitously expressed (HPA). Mutations in this gene have already been associated with the congenital disorder of glycosylation (MIM: 603147) (AR). Patients suffer from axial hypotonia, ataxia and psychomotor retardation among other symptoms. The first manifestations appear in the early childhood contrary to the affected individuals of this family; therefore, the gene was dismissed.
- *REG3A:* codifies for a C-type bactericidal lectin acting against Gram-positive bacteria. It is exclusively expressed in the gastrointestinal tract, the appendix and the skin (HPA). According to its function, this gene is unlikely to be related to the disease in fATX-166.
- *MMP16:* This gene encodes the ubiquitous homonym metalloproteinase involved in the remodeling of the extracellular matrix (HPA). It has been observed that MMP16 is expressed in microglial cells in the brain (Yoshiyama *et al.*, 1998). In HGMD Professional, a mutation in this gene has been presumably related to autism spectrum disorder, not correlating with the familiar phenotype.
- *DAB2IP:* This gene codifies for a scaffold protein involved in many signaling pathways; including immune responses, cellular apoptosis and brain development. The highest RNA expression levels are found in the cerebellum (HPA). Several mutations in this gene have been involved in schizophrenia, autism spectrum disorder and breast cancer (HGMD Professional). Both of the detected mutations present more than 8 alleles in control population (gnomAD), being probably benign according to Sherloc criteria (Nykamp *et al.*, 2017).
- *OR8G1:* encodes an olfactory receptor, which is overexpressed in testis (HPA). Variants found in this gene were considered to possibly be sequencing artifacts because four mutations were detected in a narrow region of less than 20 bp. Furthermore, c.853C>T (p.L285L) and c.861C>T (p.P287P) were detected in fATX-163. Finally, three out of four mutations present more than 8 alleles in gnomAD, being probably benign (Nykamp *et al.*, 2017).
- *ZNF676:* codifies for a protein involved in transcriptional regulation. mRNA has been detected ubiquitously having low expression in the brain (HPA). In addition, *ZNF676* variants were also considered as possible sequencing artifacts. Firstly, because the five mutations are contained in a narrow region of less than 20 bp and due to the quality of the reads. Furthermore, three of the mutations have 8 or more alleles in control population (gnomAD), being probably benign according to Sherloc criteria (Nykamp *et al.*, 2017). Therefore, these mutations can also be ruled out as disease-causing variants.

4.2 SCA36 screening

In this study, a cohort of 52 patients presenting with adult-onset ataxia was analyzed in order to determine if they were positive for SCA36, as they were lacking from a genetic diagnosis. SCA36 is an AD disorder caused by a hexanucleotide expansion (GGCCTG) located in intron 1 of the *NOP56* gene. Hence, the presence of a single expanded allele indicates SCA36.

The diagnosis is based on two different PCR reactions: firstly, a standard PCR and secondly, a RP-PCR.

4.2.1 Standard PCR, testing and fragment analysis

Standard PCR was done in order to amplify non-expanded alleles as described in 3.2.4.1 Section. Then, the obtained products were tested by electrophoresis. **Figure 7** represents the electrophoresis results from patients 40 to 47. This procedure enables to check whether the reaction was successfully completed as well as the estimate concentration of the products.

The expected amplicon length is of 174 bp (Ensemble Genome Browser 37). However, the amplified non-expanded alleles are slightly longer (**Figure 7**). This can be explained because the reference genome includes 7 hexanucleotide repeats (Section 3.2.4.4); nevertheless, it has to be taken into account that the most common allele identified in European populations is of 9 repeats (Garcia-Murias *et al.*, 2012). What is more, non-expanded alleles range from 5 to 14 repeats. Therefore, when comparing the amplified alleles, they are of similar lengths, although small differences emerge as patients may present a different number of hexanucleotide repeats.

In addition, two bands can be distinguished in samples from individuals 40, 41, 42, 44 and 45. This indicates the presence of two alleles of different lengths. By contrast, individuals 43, 46 and 47 show a single band, which suggests the presence of a single allele or two alleles of the same length. Nevertheless, these results have to be confirmed by fragment analysis, performing a higher resolution electrophoresis.

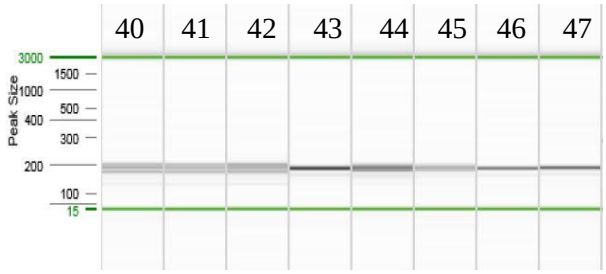

Figure 7: Standard PCR electrophoresis from patients 40 to 47. Green bands represent the molecular weight marker being the lower one of 15 bp and the upper one of 3,000 bp. The darker bands are the PCR products of varying lengths.

Following testing, fragment analysis is performed. The reagents and conditions used are described in section 3.2.4.3. The concentration of the PCR products has to be taken into account in order to prepare the samples. In this case, they were diluted 1:10 based on the amplification efficiency observed in **Figure 7**.

Figure 8 shows an example of a patient in which a single allele was detected by fragment analysis. Hence, patient 24 presents a single allele of 166 bp. Based on these observations; this patient is either homozygous or positive for the expansion, which cannot be detected with the standard PCR. Consequently, RP-PCR is necessary for confirmation.

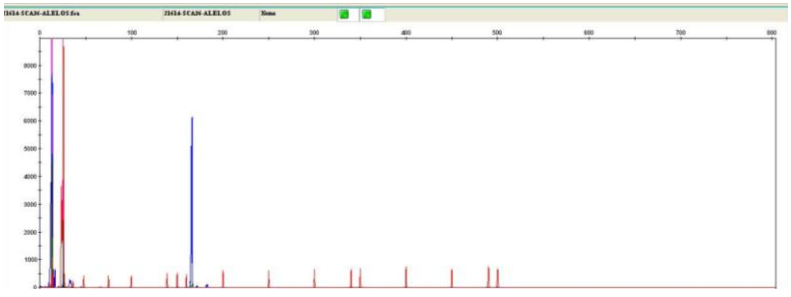

Figure 8: Patient 24 fragment analysis results displayed in GeneMapper program. A single allele of 166 bp is detected.

By contrast, **figure 9** shows an example of a patient having two non-expanded alleles of different lengths and therefore both of them can be detected by standard PCR amplification. In this case, patient 34 presents an allele of 166 bp and another of 183 bp. Consequently, this patient is negative for SCA36. Nevertheless, RP-PCR is performed anyway for confirmation.

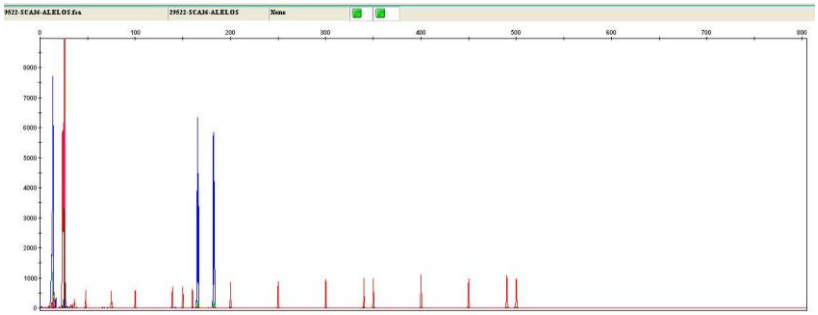

Figure 9: Patient 34 fragment analysis results displayed in GeneMapper program. Two alleles of 166 and 183 bp are detected.

4.2.2 RP-PCR and fragment analysis

RP-PCR enables for the amplification of expanded alleles and therefore for SCA36 diagnosis. The reagents and conditions used are described in Section 3.2.4.2. As previously explained, a set of products of variable lengths will be obtained. As a result, PCR products cannot be tested by conventional electrophoresis and fragment analysis is directly performed without diluting the samples.

Figure 10 shows an example of a patient positive for SCA36, having the hexanucleotide expansion. As it can be observed, a wide range of fragments of increasing length are displayed in the graph resembling a stairway. This shape is produced because the larger the expanded fragment, the lower the amplification efficiency and therefore the lower the fluorescence intensity. It must be taken into account that the number of hexanucleotide repeats cannot be determined by RP-PCR, unlike in non-expanded alleles. Other techniques such as Southern Blot can be used for this purpose.

Patient SGT-1432 presented a single allele in the former fragment analysis (standard PCR). From these observations, it can be confirmed that he is SCA36 positive and therefore, he was not homozygous for a non-expanded allele.

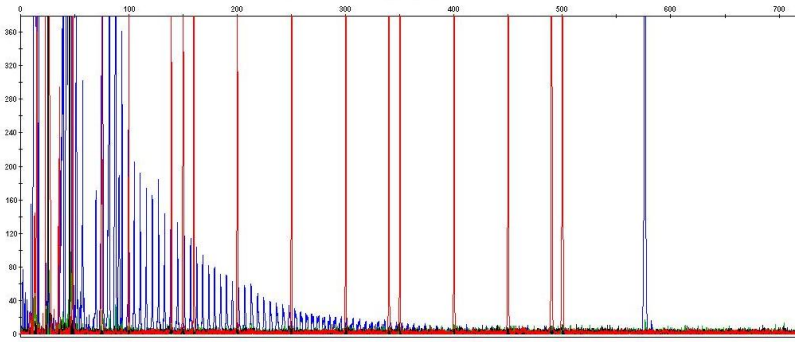

Figure 10: Patient SGT-1432 fragment analysis results displayed in GeneMapper program. A range of fragments of different lengths can be observed. The patient is positive for SCA36.

Then, fragment analysis results obtained from both PCRs were compared in order to detect inconsistencies (**Appendix VIII**).

In this case, a cohort of 52 patients with adult-onset ataxia was analyzed. Firstly, standard PCR results were analyzed and the most common identified allele was of 183 bp, having 9 repeats (Section 3.2.4.4), which is the most prevalent allele in the Spanish population (Garcia-Murias *et al.*, 2012). In this case, 35 out of 52 patients presented two alleles of different length, therefore being negative for the expansion. This result was confirmed for all of the cases by RP-PCR. By contrast, a single allele was detected by fragment analysis in 17 out of 52 patients, indicating either homozygosis or the presence of the expansion that could not be detected by standard PCR. Then, fragment analysis following RP-PCR confirmed that 10 of them were homozygotes, the remaining 7 were positive for SCA36. Positive results were confirmed twice by RP-PCR.

The analyzed 52 patients belong to 43 unrelated families. The 7 patients diagnosed with SCA36 belonged to three different families:

- **fATX-163.** The suggested pattern of inheritance derived from the family pedigree is AD (**Figure 4**), matching with the one expected from SCA36.The three individuals that undergone WES analysis were diagnosed with SCA36: SGT-1432, SGT-1433 and SGT-1459. Based on these results, other two affected individuals from the family were analyzed and confirmed.
- **Patient 25** is a 60-year-old man presenting a family history matching with AD inheritance. The initial symptoms consisted of horizontal diplopia and progressive gait instability, dysarthria and dysphagia and were firstly detected at the age of 44. His reflexes are also affected.
- **Patient 28** is a 73-year old woman who presents a family history that correlates with an AD inheritance. The first symptoms appeared at the age of 58 and where characterized by slow progression, correlating with SCA36. She suffers from pyramidal syndrome, dysarthria, nystagmus and head tremors. MRI shows cerebellar atrophy.

4.3 Analysis of the *POLR3A* c.3688G>A variant

The objective of this project relies on the determination of c.3688G>A variant effects on *POLR3A* splicing and its association with HSP in order to establish a genetic diagnose for SGT-1426. *POLR3A* encodes the largest catalytic subunit of RNA polymerase III, which catalyses the transcription of DNA into small RNAs, specifically tRNA and the 5S ribosomal subunit.

Recent studies have related compound heterozygous mutations in *POLR3A* to affected individuals suffering from CA and HSP (Minnerop *et al*., 2017). What is more, affected individuals having c.1909+22 G>A mutation in at least one allele show a unifying phenotype, characterized by a combination of gait and limb ataxia, pyramidal involvement and hyperintensities in the cerebellar peduncles (MRI) starting before the second decade of life (Infante *et al*., 2020), which correlates with the SGT-1426's clinical signs (Section 3.3).

SGT-1426 presents with sporadic HSP, constituting a challenging diagnostic procedure. To address this issue, AtxSPG-365 gene panel (**Appendix IV**) was performed to the proband, his parents and two unaffected siblings (**Figure 6**). Following gene panel data analysis, candidate mutations in *POLR3A* were identified. Firstly, SGT-1522 (unaffected father) and SGT-1426 (proband) presented c.1909+22G>A mutation in heterozygosis (**Figure 11**). As previously stated, this mutation has already been classified as pathogenic. The A allele activates a cryptic splice site located 22 bp downstream from the donor site of intron 14. As a result, 19 bp from intron 14 are inserted in the coding sequence resulting in premature stop codon and reduced levels of RPC1 protein. Nevertheless, heterozygote individuals harbouring this mutation are healthy.

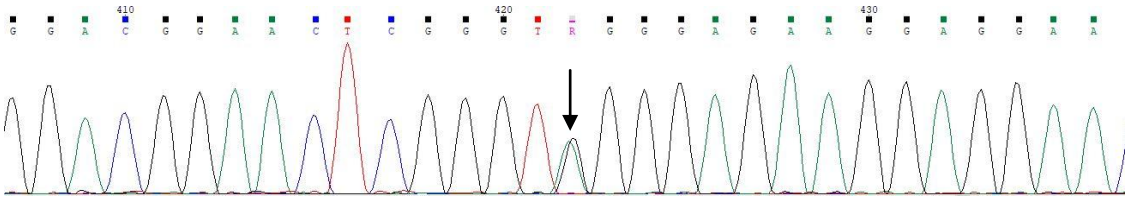

Figure 11: Electropherogram from intron 14 in SGT-1426 to detect c.1909+22G>A mutation; DNA sample was sequenced with primer reverse. Results displayed in Chromas (version 2.5).The arrow indicates the position of the mutation; A and G alleles are detected. SGT-1426 is positive for the mutation in heterozygosis.

Then, c.3688G>A missense mutation was identified in heterozygosis in four out of five individuals of the family (**Figure 6**), including the proband (**Figure 12**). Hence, SGT-1426 was the only compound heterozygote in the family presenting both variants. What is more, the

mutation is annotated in gnomAD database with a very low frequency: MAF= 3,98x10-6 (1 allele/251076 alleles). Based on this observation, the mutation is likely to be damaging according to Nykamp *et al*. (2015). Nonetheless, *in silico* and experimental approaches were required to further confirm the hypothesis.

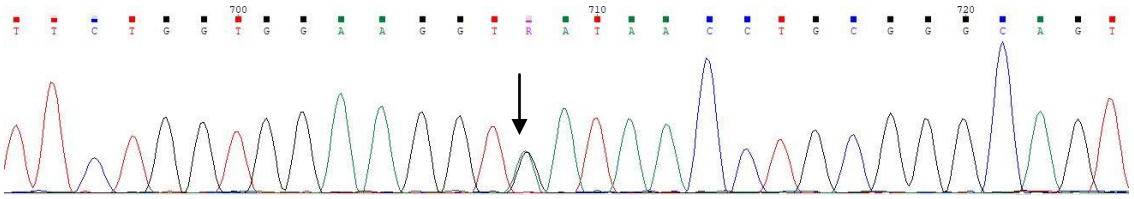

Figure 12: Electropherogram from exon 28 in SGT-1426 to detect c.3688G>A mutation; DNA sample was sequenced with primer reverse. Results displayed in Chromas (version 2.5). The arrow indicates the position of the mutation; A and G alleles are detected. SGT-1426 is positive for the mutation in heterozygosis.

4.3.1 *In silico* and experimental analysis

The missense c.3688G>A change is located in exon 28 of *POLR3A* gene. Numerous missense mutations in this gene have been described to affect the splicing; that is why, the first approach consisted of the application of *in silico* predictors. The three of them determined possible alterations of the splicing produced by c.3688G>A variant. HSF software indicated a putative anomalous splicing produced by an alteration of an exonic splice enhancer site. In addition, both NetGene and NNSplice suggested the appearance of a new donor splice site as a consequence of the mutation (**Appendix V**).

In silico analysis results suggested a presumed alteration of the splicing produced by c.3688G>A variant; nevertheless, experimental confirmation was required. Then, mRNA samples were extracted from peripheral blood of SGT-1426 (proband), compound heterozygote, and JP (healthy control), negative for both mutations.

The objective of this procedure relied on the obtaining of a cDNA sample from both individuals through a non-invasive procedure, blood extraction, as *POLR3A* is ubiquitously expressed (HPA). Then, cDNA is obtained following retrotranscription reaction (Section 3.3.1). The *POLR3A* relative expression levels in blood have to be compared with those of *GAPDH*, a housekeeping gene which is constitutively expressed whose expression levels are determined by PCR and electrophoresis. The PCR reagents and conditions are indicated in **tables 6** and **7** respectively, and the primers in **table 12**. Then, samples were tested by electrophoresis and significant differences in expression between *GAPDH* and *POLR3A* genes were not observed. In addition, this procedure serves as quality control for retrotranscription reaction.

Having both cDNA samples, from SGT-1426 and JP, a pair of primers was designed in order to amplify the target region containing the variant (exon 28). The proposed hypothesis suggests that c.3688G>A mutation leads to an alteration of the splicing; consequently, primers hybridize in exons 26 and 30 to detect possible exon skipping or other major deletions of the adjacent ones. In this case, the wild type allele is of 522 bp (Ensembl Genome Browser 37) and it is observed in both individuals, the healthy control and the proband (**Figure 13**). In contrast, SGT-1426, who has the mutation in heterozygosis, presents a second band of shorter length corresponding to the mutant allele (**Figure 13**).

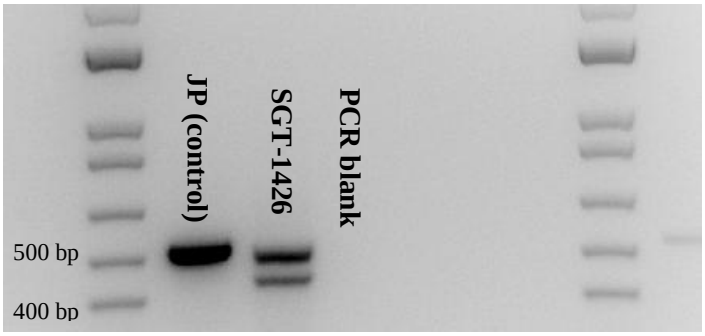

Figure 13: Testing results for *POLR3A* (cDNA) amplification to determine the c.3688G>A effects on splicing. Two DNA samples were used: JP (healthy control) and SGT-1426 (proband), a negative control was also included. The molecular weight marker employed is 1Kb Plus DNA Ladder, observed at either side of the image.

These results indicate that c.3688G>A leads to an alteration of the splicing; however, for further confirmation both samples were investigated as indicated in Section 3.1.5.2. The electropherogram obtained from the latter shows the presence of a double read in part of the sequence, which means that two different alleles are sequenced (**Figure 14**). Consequently, it can be stated that c.3688G>A mutation leads to an alteration of the splicing in *POLR3A* gene, at least *in vitro*.

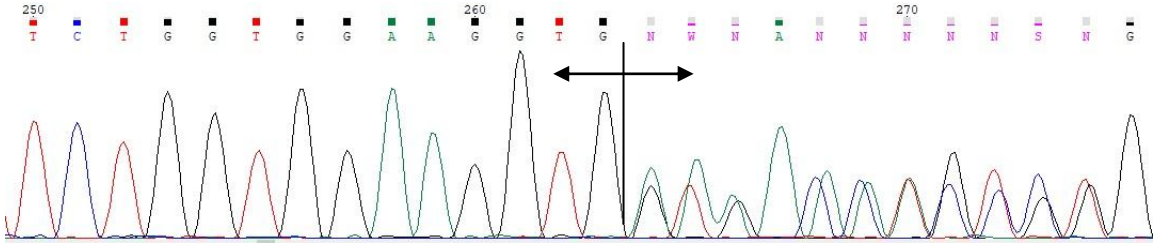

Figure 14: Electropherogram from exon 28 in SGT-1426 to detect both alleles; cDNA sample was sequenced with primer forward. Results displayed in Chromas (version 2.5). The vertical line indicates the start of the double read.

Furthermore, it was necessary to determine the molecular mechanism by which the mutation alters the splicing of *POLR3A*; nevertheless, it was suggested that c.3688G>A may lead to the activation of a cryptic splice site as c.1909+22G>A mutation (Minnerop *et al*., 2017). In this sense, the two amplified alleles in SGT-1426 have to be separated and sequenced individually. Nevertheless, based on **Figure 13**, both alleles do not differ more than 100 bp from each other, complicating their separation. In order to ensure an efficient isolation of each of the alleles, an agarose gel at 2.5% was used. Once separation was completed, each of the bands was purified from the agarose gel as indicated in Section 3.3.4, and then the two alleles were sequenced with both primers (Section 3.1.5.2).

The electropherogram of the mutant allele (**Figure 16**) presented a single read throughout the whole sequence, indicating efficient separation procedure from the wild type. In this sense, the obtained mutant allele sequence was manually compared with the reference *POLR3A* cDNA sequence (Ensembl Genome Browser 37). The results suggest that c.3688G>A mutation leads to the activation of a donor cryptic splice site adjacent to the mutation; resulting the deletion of 74 bp from exon 28. The mechanism is illustrated in **Figure 15**. As a result, the mutant allele is 448 bp long, corresponding to the size observed in **Figure 13**.

Figure 15: c.3688G>A mutation effects on splicing. Primers are indicated in yellow. Different exons are indicated in different colors and the deleted region in exon 28 is crossed out.

Nevertheless, the electropherogram of the wild-type allele presented a double read, as in **Figure 14**, indicating inefficient separation of both alleles due to the slight length difference among them. Thus, a larger separation period at lower electric potential could be advisable.

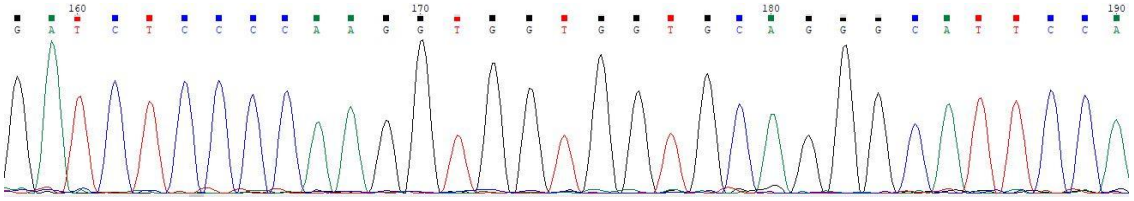

Figure 16: Electropherogram from the mutant allele of SGT-1426 sequenced with primer forward. Results displayed in Chromas (version 2.5) A single read is observed.

5. Discussion

5.1 WES in diagnosis efficiency

In this study, three families affected with adult-onset hereditary ataxia were analyzed by WES technique after resulting negative for the previous genetic tests performed (**Appendix I**). Among the different NGS approaches, WES presents the highest diagnosis efficiency when applied to a heterogeneous cohort of undiagnosed patients with HA (Galatolo *et al*., 2018).

As a result of WES analysis, an extensive list of identified variants was provided; nevertheless, it remains challenging to establish the pathogenic contribution of the mutation to the familial phenotype. Consequently, the combined results from *in silico* and experimental approaches are required. In this sense, mutations were filtered and prioritized following the pipeline described in Sections 3.1.3 and 3.1.4. Several criteria were taken into consideration in order to select candidate mutations; for instance, the familial pattern of inheritance, the MAF in gnomAD, the impact on protein functionality or the protein expression among others. Finally, candidate mutations were proposed for the three families.

In fATX-167, the c.1169T>C novel missense mutation in *MUTYH* gene was proposed as a disease-causing variant based on preliminary *in silico* analysis (Section 4.1.1). What is more, although *MUTYH* mutations have not been related to ataxia or other neurodegenerative disorders in OMIM or HGMD, several studies have related MUTYH activity to neurodegeneration under oxidative stress conditions (Sheng *et al*., 2012). Oxidative stress is a major hallmark observed in many neurodegenerative disorders such as Alzheimer (Al.D) or Parkinson's disease (PD) (Chen *et al.,* 2012), leading to an increase in 8-oxoG, an oxidized nucleotide base. In this case, MUTYH enzymatic activity repairs mispaired A by BER; nevertheless, under excessive oxidative conditions this mechanism causes increased single strand breaks, eventually generating neuronal cell apoptosis (Nakatake *et al*., 2016). Regarding *MUTYH* involvement in ataxia Shen *et al.* (2016) identified increased levels of both 8-oxoG and MUTYH protein in FA mouse model. Furthermore, *MUTYH* has been associated with cerebellar atrophy in Arabian horses, characterized by degeneration of PC and granular cells, involved in coordination of voluntary movements (Scott *et al*., 2018). This pathology is caused by a single SNP, inherited as an AR trait, located in the promoter of *MUTYH*. Consequently, it affects the degree of methylation in the promoter, the levels of isoform expression as well as the subcellular localization of MUTYH within the cerebellum of affected horses. As a result of a decrease in methylation, MYT1L transcription factor binds to *MUTYH* promoter due to loss of methylation affecting the expression of the different isoforms. Regarding their phenotype, affected horses present with analogous symptoms to the ones described in patients, consisting of uncoordinated limb movements, head tremors as well as difficulties when rising (Scott *et al.,* 2018).

In conclusion, *MUTYH* is likely to be associated with the familial phenotype, not only because it has been related with neurodegenerative disorders but specifically to cerebellar atrophy in a horses. Nevertheless, further studies are required to determine the presumed non-Mendelian pattern of inheritance observed in SGT-1457, lacking of the mutation, and functional studies will be required to confirm the pathogenesis of c.1169T>C variant.

In fATX-163 a candidate mutation in *HPCAL1* gene was identified by WES analysis. However, this family was finally diagnosed with SCA36. *HPCAL1* encodes the hippocalcin-like protein 1, also called VILIP-3 protein, which is a neuronal calcium sensor (NCS). The NCSs constitute a heterogeneous group of proteins involved in signal transduction within the central nervous system (CNS), upon rising intracellular levels of Ca^{+2}. They are primarily located in neurons and retinal photoreceptors. Visin-like proteins (VILIP), including VILIP-3, are usually located in the cytosol and undergo conformation changes upon high affinity binding with Ca^{+2}, leading to their interaction with the plasma membrane (Burgoyne, 2007). Each of the VILIP proteins presents distinct expression patterns and non-overlapping functions; concerning VILIP-3, it is

specifically expressed in the cerebellar cortex: in Purkinje cells (PCs) and in granular cells (Braunewell *et al.,* 2012). The degeneration of PCs is a typical hallmark observed in SCAs (Kasumu *et al.,* 2012). Furthermore, 17 loci have been associated with SCAs so far. Nevertheless, the pathophysiological mechanism leading to PC apoptosis and cerebellar atrophy has not been determined in many cases, preventing the development of a treatment. However, Kasumu *et al.* (2012) suggested that dysregulations in Ca^{+2}-homeostasis were the mechanism underlying PC apoptosis. Hence, given the relevance of Ca^{+2}-homeostasis in CNS signal transduction and the involvement of *HPCAL1* in such pathways; it was considered a candidate mutation. However, several observations underpinned that, regardless of its involvement in Ca^{+2} signalling, *HPCAL1* could not be involved in fATX-163 phenotype. Firstly, Groblewska *et al*. (2015) related VILIP-3 with AD in which Ca^{+2}-dysregulation is produced as a consequence of the formation of amyloid-beta plaques. Other VILIP proteins, such as VILIP-1, have been described as pathogenic factors in Alzheimer disease, amyotrophic lateral sclerosis and PD (Groblewska *et al*., 2015). Secondly, VILIP-3 remains uncharacterized and its function has not been confirmed yet, although it may be involved in the activation of MAPK based on its homology to other NCS proteins. In addition, *HPCAL1* is not associated with any disorder in OMIM, as well as VILIP-1. Consequently, although *HPCAL1* is a relevant component involved in the signal transduction of the CNS, it is probably not the causative mutation in fATX-163. Nevertheless, further co-segregation and functional studies will be required to dismiss the variant.

Finally, it was confirmed that fATX-163 was SCA36 positive and therefore, the c.275G>A variant in *HPCAL1* was ruled out. SCA36 is a form of ADCA and it correlates with the observed pattern of inheritance (**Figure 4**). It is the most common form of ADCA in Galicia due to a founder effect (Garcia-Murias *et al*., 2012). It is tempting to speculate that some of the unsolved Spanish cases in other clinical series may be explained by SCA36. Furthermore, the hexanucleotide expansion of *NOP56* gene was detected and confirmed in six different affected individuals of the family (**Appendix VIII**). What is more, the clinical manifestations associated with SCA36 correlate with patients from fATX-163 as indicated in Section 5.2. Based on these results, it can be concluded that fATX-163's patients suffer from SCA36.

In fATX-166, affected individuals were compound heterozygotes for variants c.1031A>G and c.66T>C in the *SHQ1* gene (Section 4.1.3). *SHQ1* encodes a protein involved in the formation and aggregation of H/ACA small ribonucleoproteins (RNPs). They constitute RNA-protein aggregates serving as structural subunits in the synthesis of the ribosome, spliceosome and telomerase. *SHQ1* is presumably a tumor suppressor gene as mutations have been detected in patients presenting with prostate cancer. In addition, patients suffering from X-linked dyskeratosis congenita also show increasing susceptibility to cancer because *SHQ1* interactions with the mutated protein are prevented, hindering its function (Walbott *et al.,* 2011). To date, *SHQ1* has not been involved in neurodegenerative disorders (Section 4.1.3) and in fact, its expression in the brain is reduced (HPA). In the aggregate, these observations may suggest that *SHQ1* may not be the underlying cause of the familial phenotype. In any way, it would be interesting to carry out validation and co-segregation analysis in order to definitively discard (or not) *SHQ1* as the gene responsible for disease in fATX-166.

In this project, candidate mutations were proposed for the three studied families analyzed by WES analysis. Unfortunately, this diagnostic approach did not enable to confirm the causative mutation in any case. Although Pyle *et al.* (2015) highlighted the efficiency of this technique, it is an intricate process involving filtering of mutations, validation, co-segregation and functional studies, some of which could not be performed to either confirm or dismiss the candidate mutations in fATX-167 and fATX-166. Furthermore, the difficulty in the determination of the familial pattern of inheritance, lacking of clinical information from some patients, may affect the outcome. What is more, Minnerop *et al.* (2017) stressed the relevance of non-coding variants, which cannot be identified by WES, in undiagnosed families. Recent studies have suggested the significant contribution of non-Mendelian patterns of inheritance to the existing

diagnosis gap, especially in HSP. In this sense, the combined effects of CNV, the reduced penetrance, thus, the existence of asymptomatic carriers and the effect of modifier alleles together with non-coding variants may have affected the efficiency of WES (Bis-Brewer *et al.*, 2018). Consequently, WGS approach would be more appropriate as a diagnostic tool in order to address the diagnosis gap in these neurodegenerative disorders.

5.2 SCA36 test and positive families

SCA36 is a form of ADCA and it is the most prevalent one in Galicia (Garcia-Murias *et al.*, 2012). Few cases have been diagnosed in the rest of Spain, so far, and importantly, to our knowledge, the investigated patients here included are not from Galicia. The diagnosis of hereditary ataxias, especially dominant cerebellar ones, is still a complex procedure. The most common forms are produced by polyglutamine (CAG) expansions: SCA1-3, SCA6, SCA7, SCA17 and DRLP. If the genetic test for the most common conditions is unsuccessful, other non-neurological clinical manifestations may guide further analysis with gene panels or single gene analysis. Regardless of exhaustive efforts, many patients do not achieve a genetic diagnosis years after the appearance of the first symptoms. In this sense, we aimed the development of a genetic test to efficiently detect SCA36 in cohorts of undiagnosed patients, hopefully contributing to diagnosis efficiency.

As previously explained, the diagnosis is based on the detection of the hexanucleotide expansion in intron 1 of the *NOP56* gene. The test was developed by using three positive controls (Section 3.2.2). Non-expanded (non-pathogenic alleles) are amplified by standard PCR whereas expanded (pathogenic) alleles are amplified by RP-PCR. Then the PCR products were analyzed by fragment analysis and compared in order to detect incongruities.

Three of the forty-three families screened were positive for SCA36 (Section 4.2.2). The observed clinical manifestations, together with the pattern of inheritance and the length of the most common allele correlate with SCA36 (Garcia-Murias *et al.*, 2012). Nevertheless, none of the positive patients in this cohort reported hearing loss; a clinical manifestation present in all affected individuals in the Costa da Morte (Garcia-Murias *et al.*, 2012). Therefore, it would be advisable to perform an audiometry to confirm or dismiss whether they present this trait.

Based on these results, it can be stated that, although SCA36 is not the most common form of ADCA in other regions of Spain apart from Galicia, including this diagnostic test in routine protocols will enhance diagnosis efficiency. This technique is simple, affordable and highly efficient, taking into account the complete penetrance of expanded alleles; which are between 600-2500 repeats. Nevertheless, RP-PCR does not permit to determine the length of expanded alleles, although it is not required to establish a definite diagnosis. Other techniques such as Southern Blot can be employed for this purpose. Regardless of the fact that an effective treatment has not been developed yet for most hereditary ataxias, the identification of the genetic cause underlying the disease is highly beneficial for genetic counseling. Finally, the identification of the mutation will enable patients to take part in future clinical trials targeting this condition.

5.3 Implications of c.3688G>A in HSP

SGT-1426 is a 33-year-old man diagnosed with sporadic HSP. As previously stated, the diagnosis of neurodegenerative diseases remains challenging, especially when dealing with sporadic cases. However, NGS technologies have proven to be effective in such cases. In this sense, AtxSPG-365 gene panel was applied to the proband and biallelic mutations in *POLR3A*, a gene recently related with HSP, were identified: c.1909+22G>A and c.3688G>A. Then, co-segregation studies were performed to detect those variants in other members of fATX-173 (**Figure 6**).Minnerop *et al*. (2017) determined that c.1909+22G>A is a relatively frequently mutation in patients suffering from HSP and CA. In fact, many affected subjects harbour the aforementioned mutation in heterozygosis together with a different variant, also in heterozygosis. The c.1909+22G>A mutation was found in heterozygosis in the proband and in

the SGT-1522 (unaffected father). The c.3688G>A mutation was identified in four relatives of the family, including SGT-1426 and his healthy mother. *In silico* analysis underpinned that the c.3688G>A variant may be deleterious. Importantly, this is an exonic change, but software predicted that this may be alter the proper splicing. In fact, the transcript analysis allowed us to conclude that c.3688G>A may activate a cryptic donor splice site, causing the deletion of 74 bp of exon 28 and a frame-shift. Consequently, a truncated protein is generated. In this sense, the combined effect of both mutations involves a significant reduction in RPC1 protein.

POLR3-related leukodystrophies are characterized by complete RPC1 protein expression loss, leading to alterations in myelin homeostasis and eventually hypomyelination. Nevertheless, Infante *et al.* (2020) suggested that c.1909+22G>A mutation mildly activates the cryptic splice site; hence the loss of the wild type allele is incomplete allowing for the expression of reduced RPC1 levels. The moderate, and not complete, reduction in RPC1 protein expression may affect the maintenance of the long corticoespinal axons generating a milder phenotype: HSP. As it can be observed, different mutations in *POLR3A* are associated with varying phenotypes depending on how they affect RPC1 expression and regardless of their location within the gene.

Extensive cohorts of patients presenting with CA and HSP were analyzed; and it was observed that patients harbouring c.1909+22G>A mutation present with a novel and homogeneous phenotype of adolescent-onset characterized by spastic ataxia, tremor and affection of the central sensory tracts combined with hyperintensities in the cerebellar peduncles (myelinated nerves connecting the cerebellum to the midbrain) (Infante *et al*., 2020). These clinical manifestations correlate with the ones described in the patient (Section 3.3).

Furthermore, hypomyelination is considered the unifying feature of *POLR3*-related leukodystrophies; however it has not been detected neither in the aforementioned patient cohorts (Infante *et al*., 2020) neither in SGT-1426. Therefore, Minnerop *et al.* (2017) classified this homogeneous c.1909+22G>A-related phenotype as HSP caused by *POLR3A* mutations.

Nevertheless, it is still controversial whether this homogeneous phenotype should be included in the *POLR3*-related leukodystrophy spectrum (Gauquelin *et al.,* 2018) as La Piana *et al*. (2016) described atypical forms where hypomyelination was not detected and these disorders present overlapping signs with HSP.

Regardless of their classification, it has become flagrant the relevance of NGS sequencing technologies, in this case gene panels, to identify genetic overlaps between neurodegenerative disorders. For example, *POLR3A* mutations were not even considered in patients lacking of hypomyelination (Cayami *et al.,* 2015). What is more, NGS techniques are usually employed to identify mutations located in coding regions. Nevertheless, focusing on the identification on mutations located in non-coding regions may constitute a great breakthrough when it comes to diagnosis efficiency, as it has been highlighted with the detection of c.1909+22G>A mutation.

Finally, these high-throughput techniques, together with bioinformatics analysis, have enabled to broaden our understanding regarding neurodegenerative diseases genetic and phenotype overlap; eventually increasing the diagnosis rate, as it has become noticeable in this case. However, despite of the available technology, there is a long way to go because many patients remain without molecular diagnosis.

6. Conclusions

Based on the objectives and results presented in this Final Degree Project (FDP), the following conclusions can be reached:

1. WES analysis constitutes a NGS diagnostic approach that enables to identify candidate mutations when dealing with genetically undiagnosed patients with HA. Nevertheless, it implies a complex and lengthy procedure, where numerous criteria have to be taken into account to confirm the variant association with the familial phenotype. *In silico* analysis, characterization of the mutation and the protein functionality, validation and co-segregation studies are fundamental strategies to discard non-concordant candidate mutations.

2. Putative disease-causing mutations were proposed for the three studied families: fATX-163, fATX-166 and fATX-167. Nevertheless, a definite diagnosis could not be established by WES analysis in neither of them.

3. In fATX-167, c.1169T>C novel missense variant in *MUTYH* was proposed as a candidate mutation involved in the clinical phenotype in this family, although inconsistencies were found in co-segregation studies as SGT-1457 (affected) does not present the mutation.

4. In fATX-166, compound heterozygous mutations c.1031A>G and c.66C>T in *SHQ1* were identified based on variant prioritization criteria. However, validation and co-segregation studies, and functional studies have to be performed, to confirm that *SHQ1* mutations contribute to the clinical picture in this family.

5. Regarding SCA36, an effective, affordable and simple diagnostic method has been developed to be implemented in routine diagnosis procedures. From a cohort of 43 families, three out of them were diagnosed with SCA36, including fATX-163's patients.

6. Although SCA36 is not the most common form of ADCA in other regions of Spain apart from Galicia, including this diagnostic test in routine protocols would enhance diagnoses efficiency in hereditary ataxias.

7. SGT-1426, who suffers from spastic paraplegia, carries compound heterozygote mutations in *POLR3A*: c.3688G>A and c.1909+22G>A. The c.3688G>A variant causes an aberrant transcript which is 74 bp shorter than the wild type due to the activation of a cryptic splice site located in exon 28, and the c.1909+22G>A mutation is a relatively frequent in HSP patients- Therefore, a definite genetic diagnosis was achieved for SGT.1426.

7. Bibliography

ALBERNAZ, P. L. M.; MAIA, F. Z. E.; CARMONA, S.; CAL, R. V. R. AND ZALAZAR, G. (2019). Atxia. In: *The new neurotology: A comprehensive clinical guide*. Ed:Springer, Cham: 181-191.

ARIAS, M.; QUINTÁNS, B.; GARCÍA-MURIAS, M. AND SOBRIDO, M. J. (2014) Spinocerebellar Ataxia Type 36. In Adam, M. P.; Ardinger, H. H.; Pagon, R. A.; Wallace, S. E.; Bean, L. J.; Stephens, K.; and Amemiya A. (Eds.), GeneReviews®

ASHIZAWA, T. AND XIA, G. (2016). Ataxia. *Continuum : Lifelong Learning in Neurology,* 22(4 Movement Disorders), 1208-1226.

BLACKSTONE, C. (2018). *Hereditary Spastic Paraplegia* in Handbook of Clinical Neurology. Ed: Elvesier, Amsterdam:633-652.

BIS-BREWER, D. M. AND ZÜCHNER, S. (2018). Perspectives on the genomics of HSP beyond mendelian inheritance. *Frontiers in Neurology*, 9.

BURGOYNE, R. D. (2007). Neuronal calcium sensor proteins: Generating diversity in neuronal Ca 2+ signalling. *Nature Reviews Neuroscience*, 8(3), 182-193.

BÜRK, K. (2017). Friedreich Ataxia: Current status and future prospects. *Cerebellum & Ataxias*, *4*, 4.

BERNARD, G.; CHOUERY, E.; PUTORTI, M. L.; TÉTREAULT, M.; TAKANOHASHI, A.; CAROSSO, G.; CLÉMENT, I.; BOESPFLUG-TANGUY, O.; RODRIGUEZ, D.; DELAGUE, V.; GHOCH, J. A.; JALKH, N.; DORBOZ, I.; FRIBOURG, S.; TEICHMANN, M.; MEGARBANE, A.; SCHIFFMAN, R.; VANDERVER, A. AND BRAIS, B. (2011). Mutations of polr3a encoding a catalytic subunit of rna polymerase pol iii cause a recessive hypomyelinating leukodystrophy. *American Journal of Human Genetics*, 89(3), 415-423.

BRAUNEWELL, K. H. (2012). The visinin-like proteins VILIP-1 and VILIP-3 in Alzheimer's disease—Old wine in new bottles. *Frontiers in Molecular Neuroscience*, 5.

CAYAMI, F.K.; LA PIANA, R.; VAN SPAENDONK, R.M.; NICKEL, M.; BLEY, A.; GUERRERO, K.; TRAN, L. T.; VAN DER KNAAP, M. S.; BERNARD, G. AND WOLF, N. I. (2015). POLR3A and POLR3B mutations in unclassified hypomyelination. *Neuropediatrics* 2015; 46: 221–8

CHEN, Y.-C.; WU, J.-S.; TSAI, H.-D.; HUANG, C.-Y.; CHEN, J.-J.; SUN, G. Y. AND LIN, T.-N. (2012). Peroxisome proliferator-activated receptor gamma (Ppar-γ) and neurodegenerative disorders. *Molecular Neurobiology*, 46(1), 114-124.

DE SOUZA, P. V. S.; DE REZENDE PINTO, W. B. V.; DE REZENDE BATISTELLA, G. N.; BORTHOLIN, T. AND OLIVEIRA, A. S. B. (2017). Hereditary spastic paraplegia: Clinical and genetic hallmarks. *Cerebellum* ,16(2), 525-551.

DUMAY-ODELOT, H.; DURRIEU-GAILLARD, S.; DA SILVA, D.; ROEDER, R. G. AND TEICHMANN, M. (2010). Cell growth- and differentiation-dependent regulation of RNA polymerase III transcription. *Cell Cycle*, 9(18), 3687-3699.

FOGEL, B. L.; LEE, H.; DEIGNAN, J. L.; STROM, S. P.; KANTARCI, S.; WANG, X.; QUINTERO-RIVERA, F.; VILAIN, E.; GRODY, W.W.; PERLMAN, S.; GESCHWIND, D.H. AND NELSON, S. F. (2014). Exome sequencing in the clinical diagnosis of sporadic or familial cerebellar ataxia. *JAMA Neurology*, 71(10), 1237-1246.

GALATOLO, D.; TESSA, A.; FILLA, A. AND SANTORELLI, F. M. (2018). Clinical application of next generation sequencing in hereditary spinocerebellar ataxia: Increasing the diagnostic yield and broadening the ataxia-spasticity spectrum. A retrospective analysis. *Neurogenetics*, 19(1), 1-8.

GARCÍA-MURIAS, M.; QUINTÁNS, B.; ARIAS, M.; SEIXAS, A. I.; CACHEIRO, P.; TARRÍO, R.; PARDO, J.; MILLÁN, M. J.; ARIAS-RIVAS, S.; BLANCO-ARIAS, P.; DAPENA, D.; MOREIRA, R.; RODRIGUEZ-TRELLES, F.; SEQUEIROS, J.; CARRACEDO, A.; SILVEIRA, I. AND SOBRIDO, M.J. (2012). 'Costa da

Morte' ataxia is spinocerebellar ataxia 36: Clinical and genetic characterization. *Brain*, 135(5), 1423-1435.

GAUQUELIN, L.; TÉTREAULT, M.; THIFFAULT, I.; FARROW, E.; MILLER, N.; YOO, B.; SUCHOWERSKY, O.; DUPRÉ, N.; TARNOPOLSKY, M.; BRAIS, B.; WOLF, N.; MAJEWSKI, J.; ROULEAU, G.A.; GAN-OR, Z. AND BERNARD, G. (2018). *POLR3A* variants in hereditary spastic paraplegia and ataxia. *Brain*, 141(1), e1-e1.

GROBLEWSKA, M.; MUSZYŃSKI, P.; WOJTULEWSKA-SUPRON, A.; KULCZYŃSKA-PRZYBIK, A. AND MROCZKO, B. (2015). The role of visinin-like protein-1 in the pathophysiology of alzheimer's disease. *Journal of Alzheimer's Disease*, 47(1), 17-32.

HÄGGLUND, M. G. A.; SREEDHARAN, S.; NILSSON, V. C. O.; SHAIK, J. H. A.; ALMKVIST, I. M.; BÄCKLIN, S.; WRANGE, O. AND FREDRIKSSON, R. (2011). Identification of SLC38A7 (Snat7) protein as a glutamine transporter expressed in neurons. *The Journal of Biological Chemistry*, 286(23), 20500-20511.

HÉBERT, S. S., AND DE STROOPER, B. (2009). Alterations of the microRNA network cause neurodegenerative disease. *Trends in Neurosciences*, 32(4), 199-206.

IKEDA, Y.; OHTA, Y.; KOBAYASHI, H.; OKAMOTO, M.; TAKAMATSU, K.; OTA, T.; MANABE, Y.; OKAMOTO, K.; KOZUMI, A. AND ABE, K. (2012). Clinical features of SCA36: A novel spinocerebellar ataxia with motor neuron involvement (Asidan). *Neurology*, 79(4), 333-341.

INFANTE, J.; SERRANO-CÁRDENAS, K. M.; CORRAL-JUAN, M.; FARRÉ, X.; SÁNCHEZ, I.; DE LUCAS, E. M.; GARCÍA, A.; MARTÍN-GURPEGUI, J.L.; BERCIANO, J. AND MATILLA-DUEÑAS, A. (2020). POLR3A-related spastic ataxia: New mutations and a look into the phenotype. *Journal of Neurology*, 267(2), 324-330.

JAYADEV, S. AND BIRD, T. D. (2013). Hereditary ataxias: Overview. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 15(9), 673-683.

KARA, E.; TUCCI, A.;MANZONI, C.; LYNCH, D. S.; ELPIDOROU, M.; BETTENCOURT, C.; CHELBAN, V.; MANOLE, A.; PITTMAN, S. A.; JAUNMUKTANE, Z.; BRADNER, S.; XIROMERISIOU, G.; WIENTHOFF, S.; SCHOTTLAENDER, L.; PROUKAKIS, C.; MORRIS, H.; WARNER, T.; BHATIA, K. P.; KORLIPARA, L. V. P.; SINGENTON, A. B.; HARDY, J.; WOOD, N. W.; LEWIS, P.A. AND HOULDEN, H. (2016). Genetic and phenotypic characterization of complex hereditary spastic paraplegia. *Brain: A Journal of Neurology*, 139(Pt 7), 1904-1918.

KASUMU, A. AND BEZPROZVANNY, I. (2012). Deranged calcium signaling in purkinje cells and pathogenesis in spinocerebellar ataxia 2 (Sca2) and other ataxias. *The Cerebellum*, 11(3), 630-639.

KOBAYASHI, H.; ABE, K.; MATSUURA, T.; IKEDA, Y.; HITOMI, T.; AKECHI, Y.; HABU, T.; LIU, W.; OKUDA, H. AND KOIZUMI, A. (2011). Expansion of intronic GGCCTG hexanucleotide repeat in NOP56 causes SCA36, a type of spinocerebellar ataxia accompanied by motor neuron involvement. *American Journal of Human Genetics*, 89(1), 121-130.

KOBAYASHI, H.; ABE, K.; MATSUURA, T.; IKEDA, Y.; HITOMI, T.; AKECHI, Y.; HABU, T.; LIU, W.; OKUDA, H. AND KOIZUMI, A. (2011). Expansion of intronic ggcctg hexanucleotide repeat in nop56 causes sca36, a type of spinocerebellar ataxia accompanied by motor neuron involvement. *The American Journal of Human Genetics*, 89(1), 121-130.

MARCHUK, D. S.; CROOKS, K.; STRANDE, N.; KAISER-ROGERS, K.; MILKO, L. V.; BRANDT, A.; ARREOLA, A.; TILEY, C. R.; BIZON, C.; VORA, N. L.; WILHELMSEN, K. C.; EVANS, J. P. AND BERG, J. S. (2018). Increasing the diagnostic yield of exome sequencing by copy number variant analysis. *Plos One*, 13(12), e0209185.

MATSUZONO, K.; IMAMURA, K.; MURAKAMI, N.; TSUKITA, K.; YAMAMOTO, T.; IZUMI, Y.; KAJI, R.; OHTA, Y.; YAMASHITA, T.; ABE, K. AND INOUE, H. (2017). Antisense oligonucleotides reduce rna foci in spinocerebellar ataxia 36 patient ipscs. *Molecular Therapy. Nucleic Acids*, 8, 211-219.

MINNEROP, M.; KURZWELLY, D.; WAGNER, H.; SOEHN, A. S.; REICHBAUER, J.; TAO, F.; RATTAY, T.W.; PEITZ, M.; REHBACH, K.; GIORGETTI, A.; PYLE, A.; THIELE, H.; ALTMÜLLER, J.; TIMMANN, D.; KARACA, I.; LENNARZ, M.; BAETS, J.; HENGEL, H.; SYNOFZIK, M.; ATASU, B.; FEELY, S.; KENNERSON, M.; STENDEL, C.; LINDIG, T.; GONZALEZ, M.A.; STIRNBERG, R.; STURM, M.; ROESKE, S.; JUNG, J.; BAUER, P.; LOHMANN, E.; HERMS, S.; HEILMANN-HEIMBACH, S.; NICHOLSON, G.; MAHANJAH, M.; SHARKIA, R.; CARLONI, P.; BRÜSTLE, O.; KLOPSTOCK, T.; MATHEWS, K.D.; SHY, M. E.; DE JONGHE, P.; CHINNERY, P.F.; HORVATH, R.; KOHLHASE, J.; SCHMITT, I.; WOLF, M.; GRESCHUS, S.; AMUNTS, K.; MAIER, W.; SCHÖLS, L.; NÜRNBERG, P.; ZUCHNER, S.; KLOCKGETHER, T.; RAMIREZ, A. AND SCHÜLE, R. (2017). Hypomorphic mutations in POLR3A are a frequent cause of sporadic and recessive spastic ataxia. *Brain*, 140(6), 1561-1578.

MOREIRA, R.; RODRÍGUEZ-TRELLES, F.; SEQUEIROS, J.; CARRACEDO, A.; SILVEIRA, I. AND SOBRIDO, M. J. (2012). 'Costa da Morte' ataxia is spinocerebellar ataxia 36: Clinical and genetic characterization. *Brain: A Journal of Neurology*, 135(5), 1423-1435.

NAKATAKE, S.; MURAKAMI, Y.; IKEDA, Y.; MORIOKA, N.; TACHIBANA, T.; FUJIWARA, K.; YOSHIDA, N.; NOTOMI, S.; HISATOMI, T.; YOSHIDA, S.; ISHIBASHI, T.; NAKANEPPU, Y. AND SONODA, K.-H. (2016). MUTYH promotes oxidative microglial activation and inherited retinal degeneration. JCI Insight, 1(15).

NÉMETH, A. H.; KWASNIEWSKA, A. C.; LISE, S.; PAROLIN SCHNEKENBERG, R.; BECKER, E. B. E.; BERA, K. D.; SHANKS, M. E.; GREGORY, L.; BUCK, D.; CADER, M. Z.; TALBOT, K.; DE SILVA, R.; FLETCHER, N.; HASTINGS, R.; JAYAWANT, S.; MORRISON, P. J.; WORTH, P.; TAYLOR, M.; TOLMIE, J.; O'REGAN, M.; VALENTINE, R.; PACKHAM, E.; EVANS, J.; SELLER, A. AND RAGOUSSIS, J. (2013). Next generation sequencing for molecular diagnosis of neurological disorders using ataxias as a model. *Brain: A Journal of Neurology*, 136(Pt 10), 3106-3118.

NYKAMP, K.; ANDERSON, M.; POWERS, M.; GARCIA, J.; HERRERA, B.; HO, Y.-Y.; KOBAYASHI, Y.; PATIL, N.; THUSBERG, J.; WESTBROOK, M. AND TOPPER, S. (2017). Sherloc: A comprehensive refinement of the ACMG-AMP variant classification criteria. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 19(10), 1105-1117.

PALAU, F., AND ESPINÓS, C. (2006). Autosomal recessive cerebellar ataxias. *Orphanet Journal of Rare Diseases*, 1, 47.

PARODI, L.; COARELLI, G.; STEVANIN, G.; BRICE, A. AND DURR, A. (2018). Hereditary ataxias and paraparesias: Clinical and genetic update. *Current Opinion in Neurology*, 31(4), 462–471.

PFEFFER, G.; PYLE, A.; GRIFFIN, H.; MILLER, J.; WILSON, V.; TURNBULL, L.; FAWCETT, K.; SIMS, D.; EGLON, G.; HADJIVASSILIOU, M.; HORVATH, R.; NÉMETH, A. AND CHINNERY, P. F. (2015). SPG7 mutations are a common cause of undiagnosed ataxia. *Neurology*, 84(11), 1174-1176.

PIANA, R. L.; CAYAMI, F. K.; TRAN, L. T.; GUERRERO, K.; SPAENDONK, R. VAN; ÕUNAP, K.; SANDER, P.; HAAK, T.; WASSMER, E.; TIMMAN, D.; MIERZEWSKA, H.; BWEE, T.; PATEL, C.; COX, H.; ATIK, T.; ONAY, H.; OZKINAY, F.; VANDERVER, A.; VAN DER KNAAP, M.J.; WOLF., N.I. AND BERNARD, G. (2016). Diffuse hypomyelination is not obligate for POLR3-related disorders. *Neurology*, 86(17), 1622-1626.

PYLE, A.; SMERTENKO, T.; BARGIELA, D.; GRIFFIN, H.; DUFF, J.; APPLETON, M.; DOUROUDIS, K.; PFEFFER, G.; SANTIBANEZ-KOREF, M.; EGLON, G.; YU-WAI-MAN, P.; RAMESH, V.; HORVATH, R. AND CHINNERY, P.F. (2015). Exome sequencing in undiagnosed inherited and sporadic ataxias. *Brain. A Journal of Neurology*, 138(2), 276-283.

RUANO, L.; MELO, C.; SILVA, M. C., AND COUTINHO, P. (2014). The global epidemiology of hereditary ataxia and spastic paraplegia: A systematic review of prevalence studies. *Neuroepidemiology*, 42(3), 174-183.

SALINAS, S.; PROUKAKIS, C.; CROSBY, A. AND WARNER, T. T. (2008). Hereditary spastic paraplegia: Clinical features and pathogenetic mechanisms. *The Lancet Neurology*, 7(12), 1127-1138.

SCHEPER, G. C.; VAN DER KLOK, T.; VAN ANDEL, R. J.; VAN BERKEL, C. G. M.; SISSLER, M.; SMET, J.; MURAVINA, T. I.; SERKOV, S. V.; UZIEL, G.; BUGIANI, M.; SCHIFFMANN, R.; KRÄGELOH-MANN, I.; SMEITINK, J. A. M.; FLORENTZ, C.; VAN COSTER, R.; PRONK C. J. AND VAN DER KNAAP, M. S. (2007). Mitochondrial aspartyl-tRNA synthetase deficiency causes leukoencephalopathy with brain stem and spinal cord involvement and lactate elevation. *Nature Genetics*, 39(4), 534-539.

SCOTT, E. Y.; WOOLARD, K. D.; FINNO, C. J.; PENEDO, M. C. T. AND MURRAY, J. D. (2018). Variation in mutyh expression in arabian horses with cerebellar abiotrophy. *Brain Research*, 1678, 330-336.

SEIDEL, K.; SISWANTO, S.; BRUNT, E. R. P.; DEN DUNNEN, W.; KORF, H.-W.AND RÜB, U. (2012). Brain pathology of spinocerebellar ataxias. *Acta Neuropathologica*, 124(1), 1-21.

SHAKKOTTAI, V. G. AND FOGEL, B. L. (2013). Clinical neurogenetics: Autosomal dominant spinocerebellar ataxia. *Neurologic Clinics*, 31(4), 987-1007.

SHEN, Y.; MCMACKIN, M. Z.; SHAN, Y.; RAETZ, A.; DAVID, S. AND CORTOPASSI, G. (2016). Frataxin deficiency promotes excess microglial DNA damage and inflammation that is rescued by PJ34. *PLOS ONE*, 11(3), e0151026.

SHENG, Z.; OKA, S.; TSUCHIMOTO, D.; ABOLHASSANI, N.; NOMARU, H.; SAKUMI, K.; YAMADA, H. AND NAKABEPPU, Y. (2012). 8-Oxoguanine causes neurodegeneration during MUTYH-mediated DNA base excision repair. *The Journal of Clinical Investigation*, 122(12), 4344-4361.

SHRIBMAN, S.; REID, E.; CROSBY, A. H.; HOULDEN, H. AND WARNER, T. T. (2019). Hereditary spastic paraplegia: From diagnosis to emerging therapeutic approaches. *The Lancet Neurology*, 18(12), 1136-1146.

SOLOWSKA, J. M. AND BAAS, P. W. (2015). Hereditary spastic paraplegia SPG4: What is known and not known about the disease. *Brain: A Journal of Neurology*, 138(Pt 9), 2471-2484.

SYNOFZIK, M., AND SCHÜLE, R. (2017). Overcoming the divide between ataxias and spastic paraplegias: Shared phenotypes, genes, and pathways. *Movement Disorders: Official Journal of the Movement Disorder Society*, 32(3), 332-345.

YANG, Z.; WANG, J.; HUANG, L.; LILLEY, D. M. J. AND YE, K. (2020). Functional organization of box C/D RNA-guided RNA methyltransferase. *Nucleic Acids Research*, 48(9), 5094-5105.

YOSHIYAMA, Y.; SATO, H.; SEIKI, M.; SHINAGAWA, A.; TAKAHASHI, M. AND YAMADA, T. (1998). Expression of the membrane-type 3 matrix metalloproteinase (Mt3-mmp) in human brain tissues. *Acta Neuropathologica*, 96(4), 347-350.

WALBOTT, H.; MACHADO-PINILLA, R.; LIGER, D.; BLAUD, M.; RÉTY, S.; GROZDANOV, P. N.; GODIN, K.; VAN TILBEURG, H.; VARANI, G.; MEIER, U.T. AND LEULLIOT, N. (2011). The H/ACA RNP assembly factor SHQ1 functions as an rna mimic. *Genes & Development*, 25(22), 2398-2408.

WELNIARZ, Q.; DUSART, I.; AND ROZE, E. (2017). The corticospinal tract: Evolution, development, and human disorders. *Developmental Neurobiology*, 77(7), 810-829.

WOLF, N.I.; VANDERVER, A.; VAN SPAENDONK, R. M.; SCHIFFMANN, R.; BRAIS, B.; BUGIANI, M.; SISTERMANS, E.; CATSMAN-BERREVOETS, C.; KROS, J. M.; PINTO, P. S.; POHL, D.; TIRUPATHI, S.; STRØMME, P.; DE GRAUW, T.; FRIBOURG, S.; DEMOS, M.; PIZZINO, A.; NAIDU, S.; GUERRERO, K.; VAN DER KNAAP, M.S. AND BERNARD, G.(2014) Clinical spectrum of 4H leukodystrophy caused by POLR3A and POLR3B mutations. *Neurology* 83:1898–1905.