UNIVERSITÀ DEGLI STUDI FIRENZE

TUSCAN Ph.D. IN NEUROSCIENCES

Cycle XXXII

COORDINATOR Prof. Renato Corradetti

An integrated, next-generation approach to identify new genes and new pathways in hereditary ataxias

Scientific Disciplinary Sector MED/39

Ph.D. Student

Dr. Daniele Galatolo

Tutor

Prof. Filippo Maria Santorelli

Tutor

Prof. Giovanni Cioni

Coordinator

Prof. Renato Corradetti

Years 2016/2019

Questa tesi è dedicata
ai miei genitori
Manuela e Stefano

Rifugi nella tempesta,
sentieri nella foresta,
stelle nella notte.

Supervised by:

Prof. Filippo Maria Santorelli

Molecular Medicine for Neurodegenerative and Neuromuscular Diseases Unit

IRCCS Fondazione Stella Maris, Pisa - Italy.

Prof. Giovanni Cioni

Department of Developmental Neuroscience

IRCCS Fondazione Stella Maris, Pisa – Italy

Department of Clinical and Experimental Medicine

University of Pisa, Pisa – Italy

Reviewed by:

Prof. Matthis Synofzik

Department of Neurodegenerative Diseases, Hertie-Institut für klinische Hirnforschung (HIH)

German Research Center for Neurodegenerative Diseases (DZNE)

University of Tübingen, Tübingen – Germany

Prof. Giovanni Stevanin

Basic to Translational Neurogenetics team, Institut du Cerveau et de la Moelle épinière (ICM)

Neurogenetics team, Ecole Pratique des Hautes Etudes (EPHE)

Sorbonne University, Paris - France

List of original publications

This Ph.D. thesis work has been composed on the basis of the following original papers and other unpublished material in preparation:

- **<u>Galatolo D</u>**, Tessa A, Filla A, Santorelli FM. 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 (2018).

- De Michele G, Lieto M, **<u>Galatolo D</u>**, Salvatore E, Cocozza S, Barghigiani M, Tessa A, Baldacci J, Pappatà S, Filla A, De Michele G, Santorelli FM. Spinocerebellar ataxia 48 presenting with ataxia associated with cognitive, psychiatric, and extrapyramidal features: A report of two Italian families. *Parkinsonism Relat Disord*. 65:91-96 (2019).

- Lieto M, Riso V, **<u>Galatolo D</u>**, De Michele G, Rossi S, Barghigiani M, Cocozza S, Pontillo G, Trovato R, Saccà F, Salvatore E, Tessa A, Filla A, Santorelli FM, De Michele G, Silvestri G. The complex phenotype of spinocerebellar ataxia type 48 (SCA48) in eight unrelated Italian families. *Eur J Neurol.* [Epub ahead of print] (2019).

- Galosi S, Barca B, Carrozzo R, Schirinzi T, Quinzii CM, Lieto M, Vasco G, Zanni G, Di Nottia M, **<u>Galatolo D</u>**, Filla A, Bertini E, Santorelli FM, Leuzzi V, Haas R, Hirano M, Friedman J. Dystonia-Ataxia with early handwriting deterioration in *COQ8A* mutation carriers: A case series and literature review. *Parkinsonism Relat Disord.* 68:8-16 (2019).

- Lieto M, **<u>Galatolo D</u>**, Roca A, Cocozza S, Pontillo G, Fico T, Pane C, Saccà F, De Michele G, Santorelli FM, Filla A. Overt hypogonadism may be not a sentinel sign of *RNF216*: two novel mutations associated with ataxia, chorea and fertility. *Mov Disord Clin Pract.* 6(8):724-726 (2019).

- Lieto M, **<u>Galatolo D</u>**, Santorelli FM, Saccà F, De Michele G, Filla A. Progressive myoclonic ataxia associated with *ATP13A2* mutation: widening Kufor-Rakeb phenotype? Submitted

- **<u>Galatolo D</u>**, Kuo ME, Mullen P, Meyer-Schuman R, Doccini S, Battini R, Lieto M, Tessa A, Filla A, Francklyn C, Antonellis A, Santorelli FM. Bi-allelic mutations in *HARS* cause a multi-system syndrome that includes microcephaly, intellectual, disability, skeletal deformities, and ataxia. Submitted

Abstract

The Hereditary ataxias (HAs) are a group of heterogenous neurological disorders associated with multiple genetic etiologies and encompassing a wide spectrum of phenotypes, where ataxia is the prominent feature. HAs are characterized by degeneration of Purkinje cell and/or spinocerebellar connections, often associated with defects in additional brain structures, and all patterns of inheritance may occur. Similar to other fields of medical genetics, Next Generation Sequencing (NGS) has entered the HA scenario widening our genetic and clinical knowledge of this condition, but routine NGS applications still miss genetic diagnosis in about two third of patients.

In this doctoral study, we applied multi-gene panels to define the molecular basis in 259 patients with a clinical diagnosis of HA and negative to tests for pathological expansion in SCA1, 2, 3, 6, 7, 8, 12, 17 and *FXN*. We found a positive molecular diagnosis in 25% of patients, whereas a similar number of patients had an uncertain diagnosis due to the presence of either variants of uncertain significance or lack of biological samples to determine segregation among family members. Hence despite a higher positive diagnostic rate compared to similar studies described in literature, a half of patients lacked any indication of the genetic cause of their disease. Using exome sequencing as a second-tier approach in some families, refractory to multi-gene panel analysis, did not significantly improved our diagnostic yield. On the other hand, NGS analysis in our cohort indicated that familial cases were more easily diagnosed rather than sporadic cases, and also that combining massive sequencing with detailed clinical information and family studies increases the likelihood to reach a molecular diagnosis. Among positive patients, we could expand clinical and allelic information in a subgroup of genes offering original description of new mutations and corroborating genetic findings with functional investigations that took advantage of different *in vitro* or *in vivo* platforms. In particular, through functional studies in *SPG7* knock-down models of *Drosophila melanogaster*, we remarked that *SPG7*, whose mutations cause spastic paraplegia type 7, has a critical role in neurons more than in skeletal muscle. The high frequency of p.Ala510Val mutation in *SPG7* observed in our cohort as well in similar studies performed elsewhere moved us to develop a humanized knock-in fruit fly model harboring that specific mutation and prepare preliminary characterizations. Similar studies in fruit fly were performed silencing *AFG3L2*, the gene causing SPAX5 in a child in association with an unusual, relatively milder phenotype. Furthermore, combination of skin fibroblasts and *Saccharomyces cerevisiae* as models was employed in the genetic characterization of new mutations in a novel recessive *HARS*-related phenotype whereas primary human cells, yeast and *Danio rerio* models were used to functionally characterize new HA-related mutations in *COQ4*.

Finally, we could expand the clinical presentation of rare causes of HAs describing new dominant mutations in *STUB1* and biallelic variants in *RFN216*, *COQ8A*, and *ATP13A2*.

Altogether, studies performed during this doctoral work further underlined the usefulness of NGS in HAs and highlighted how NGS technologies rely on the integrated use of family and clinical studies and different *in vitro*/*in vivo* platforms to substantiate molecular findings. The latter platform will be also a tool for future investigations to dissect pathogenesis and to improve therapies.

Index

1. Introduction

1.1 Hereditary ataxias (HAs)

The Hereditary ataxias (HAs) are a group of heterogeneous neurological disorders encompassing a wide spectrum of phenotypes, where ataxia (literally "absence of order") is the prominent feature. Gait abnormalities, lack of coordination, dysarthria and dysmetria are the most common traits of HA patients (Jayadev and Bird, 2013). HAs are characterized by degeneration of Purkinje cell and/or spinocerebellar connections, often combined with atrophy of other regions of both the central and peripheral nervous systems.

Although there is a scholastic classification based on pattern of transmission and disease-gene relationship, examples of overlaps between clinical syndromes are growing at a rapid pace and it is common to see HA overlapping with other neurological diseases (e.g. hereditary spastic paraplegias, epilepsies, choreiform disorders, parkinsonism) (Synofzik and Schule, 2017). Considering the many genes playing a disease role in several neurologic conditions and the overlap in pathogenic mechanisms, it seems more appropriate to consider ataxia disorders as part of a broader spectrum of phenotypes ranging from HA to more complex syndromes such as HSP. Yet, in clinical practice the scholastic classification in autosomal dominant spinocerebellar ataxias (SCAs) and autosomal recessive spinocerebellar ataxias (SCARs) remains useful to orient molecular tests.

X-linked ataxias are very uncommon, and ataxia in disorders caused by mutations in mitochondrial DNA (mtDNA) is frequent but only as symptom in very complex syndromes.

1.1.1 Autosomal dominant spinocerebellar ataxias (SCAs)

Autosomal dominant forms of HAs are represented by spinocerebellar ataxias (SCAs), that are more than 40 clinical conditions (Klockgether et al., 2019) classified with progressive numbers based on the time of gene cloning (Didonna and Opal, 2016). SCAs have an overall estimated prevalence of $1.5\text{-}4.0\times10^{-5}$ (Ruano et al., 2014), though these figures differ in specific populations mainly because of founder effects. Clinical pictures of dominant HAs are generally more homogeneous than recessive HAs. Adult or late onset forms are usual, but early onset also may occur (Durr, 2010). Progressive loss of balance and coordination with dysarthria are the most common signs in SCAs patients, but each SCA subtype has usually variable and peculiar features. Indeed, even is in part obsolete, Harding's classification of dominant HAs (Harding, 1982) in

different clinical subgroups (ADCA I, ADCAII and ADCA III) remains a useful milestone and it is still used by clinicians.

Most common SCAs are caused by pathologic trinucleotide repeat expansions in coding regions (SCA1 MIM# 164400, SCA2 MIM# 183090, SCA3 MIM# 109150, SCA6 MIM# 183086, SCA7 MIM# 164500, SCA17 MIM# 607136, DRPLA MIM# 125370), in untranslated regions (SCA8 MIM# 608768, SCA12 MIM# 604326) or in intronic sequence (SCA10 MIM# 603516). Penta or hexanucleotide repeat expansion also occur (SCA31 MIM# 117210, SCA36 MIM# 614153, SCA37 MIM# 615945). Other SCAs are due to conventional mutations. The positive correlation between the size enlargement of trinucleotide expansions and earlier and more severe symptoms (anticipation) and the possibility that some affected individuals may not show the disease even when harboring the disease-causing allele (incomplete penetrance) are common features in SCAs. Because of the instability of trinucleotide expansion, age of onset of the disease may have an intrafamilial variability in polyglutamine expansion SCAs, that are also progressive and characterized by a higher severity of the early-onset cases rather than late-onset cases. These types of SCAs are also outlined by a multisystemic clinical picture and a widespread neurodegeneration. On the other hand, conventional mutation SCAs has yet variable age of onset but more often in childhood, they are usually slowly progressive, and their earlier onset is not linked to a more severe phenotype. Moreover, cerebellar features are predominant, as confirmed by the typical pure cerebellar atrophy (Durr, 2010).

Possible mechanisms encompassing the cause of SCAs can be grouped in four main subgroups (Figure 1): i) Polyglutamine expansions that contribute to pathogenesis through altered conformation, interactions and oligomerization of proteins establishing abnormal aggregates, and loss of nuclear integrity caused by intranuclear interactions, transcription dysregulation and DNA damage ii) Dysfunctions in ionic channels and transporters iii) Impairments in proteins involved in signal transduction pathways iv) Toxic activity of aberrant RNA caused by large non-coding repeat expansions or by repeat associated non-ATG mediated (RAN) translation (Klockgether et al., 2019).

Despite the enormous progresses in broadening the knowledge in SCAs, no therapies are actually available, but recent and still ongoing randomized clinical trials offer promising results for symptomatic treatments (Schniepp et al., 2012) (Romano et al., 2015) (Lei et al., 2016) and disease-modifying therapies using gene silencing (Keiser et al., 2016) (Scoles and Pulst, 2018). New opportunities for gene therapies are also emerging (Coarelli et al., 2018).
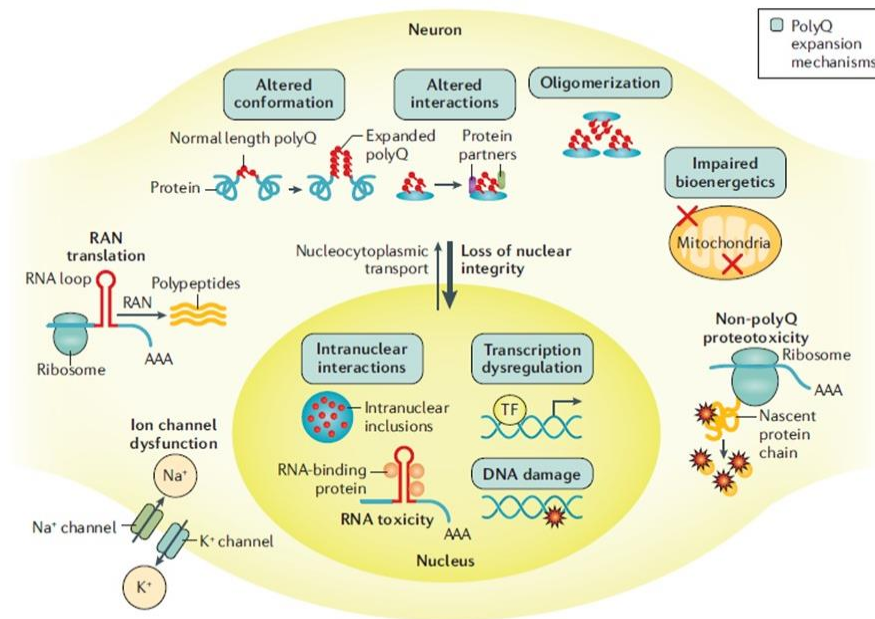
Figure 1. Common disease mechanisms underlying dominant spinocerebellar ataxias (SCAs). Processes involved in SCAs etiopathology can be summarized in protein aggregation and disfunction caused by polyglutamine expansions, impairments in channels and transporters, abnormalities in signal transduction pathways, and toxic activity of aberrant RNA. (from Klockgether et al., 2019).

1.1.2 Autosomal recessive spinocerebellar ataxias (SCARs)

Recessive forms of HAs (SCARs) are genetically and clinically more heterogeneous than SCAs, with a prevalence of 3-6/100.000 (Synofzik and Nemeth, 2018). Generally, progressive cerebellar degeneration is a common feature, whilst degenerative processes comprise other nervous system structures but also other tissues, often leading to complex and multisystemic disorders characterized by inter- and intra-familial variability and usually with onset in infancy or childhood. More than 100 genes are so far involved in SCARs (Synofzik and Nemeth, 2018), but phenotypes are usually combined with spasticity, epilepsy and neuropathy (Anheim et al., 2012) (Synofzik and Schule, 2017), whilst about an equal number of genes is cause of recessive diseases where ataxia is part of the clinical picture.

Contrary to SCAs, sporadic cases are more frequent and with more severe and disabling features. Friedreich ataxia (FA, MIM# 229300), caused by large intronic GAA expansion in *FXN* (Campuzano et al., 1996), is by far the most common recessive ataxia, since it is the cause of about 25% of all SCARs and up to 50% of SCARs patients in Europe. A similar frequency is observed in Cerebellar Ataxia, Neuronopathy, Vestibular Areflexia Syndrome (CANVAS, MIM# 614575), due

to intronic AAGGG repeat expansion in *RFC1* (Cortese et al., 2019) yet characterized by late age of onset. Interestingly, Autosomal Recessive Spastic Ataxia of Charlevoix-Saguenay (ARSACS, MIM# 270550) and SCAR8 (MIM# 610743), caused by mutations in *SACS* and *SYNE1*, respectively were originally considered to be restricted to Quebecois patients of French Canadian descent due to founder effect, but they now represent among the most frequent cause of recessive HAs worldwide (Synofzik et al., 2013) (Fogel et al., 2014) (Synofzik et al., 2016) (Coutelier et al., 2018) (Synofzik et al., 2019). Also *SPG7* (SPG7, MIM# 607259), originally considered a cause of hereditary spastic paraparesis (HSP) (Casari et al., 1998), was found to be a common cause of recessive HAs or spastic-ataxias (Van Gassen et al., 2012), and to date is the most frequent one after FA (Choquet et al., 2016) (Hewamadduma et al., 2018) (Mancini et al., 2018) (Coarelli et al., 2019). *APTX* (Ataxia with oculomotor apraxia 1, MIM# 208920), *SETX* (Ataxia with oculomotor apraxia 2, MIM# 606002), and *ATM* (Ataxia-telangiectasia, MIM# 208900) are also frequent, each representing 2-10% of all SCARs (Synofzik and Nemeth, 2018).
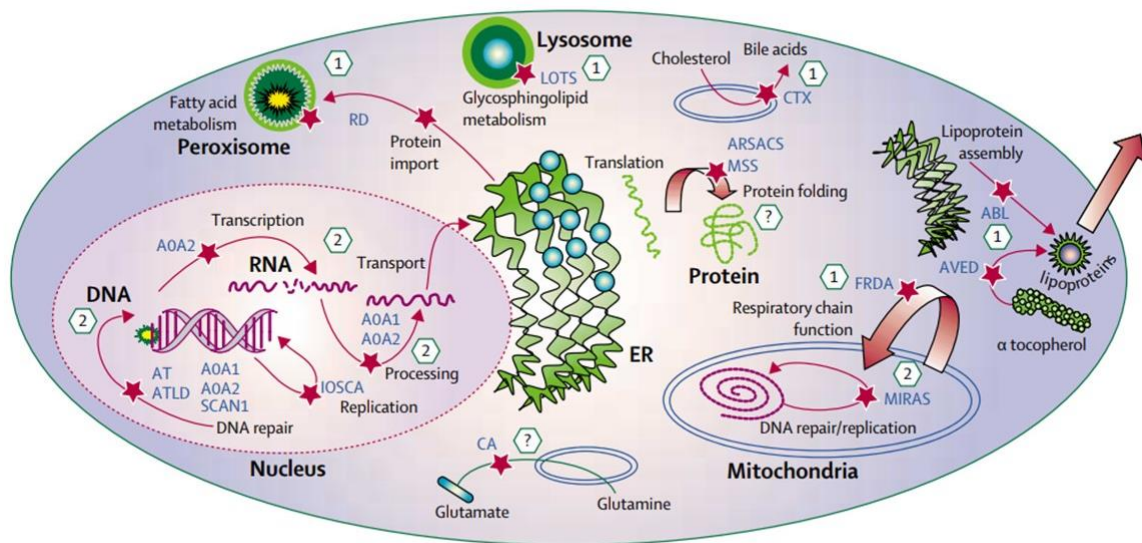
Figure 2. Molecular pathogenesis of the autosomal recessive ataxias (SCARs). Spinocerebellar Autosomal Recessive Ataxias (SCARs) are mainly caused by defects in mitochondrial processes, DNA repair/genome stability, and lipid metabolism. Pathways involved are indicated by a red star. (?) remarks pathogenesis still unclear, (1) denotes abnormalities in metabolic processes, and (2) specifies impairments in genetic stability. FRDA= Friedreich's ataxia. AVED= ataxia with vitamin E deficiency. ABL= abetalipoproteinaemia. RD= Refsum's disease. LOTS= late-onset Tay-Sachs disease. CTX= cerebrotendinous xanthomatosis. MIRAS= mitochondrial recessive ataxia syndrome. SCAN1= spinocerebellar ataxia with axonal neuropathy. AT= ataxia telangiectasia. AOA1= ataxia with oculomotor apraxia, type 1. AOA2= ataxia with oculomotor apraxia, type 2. ARSACS= autosomal recessive ataxia of Charlevoix-Saguenay. IOSCA= infantile-onset spinocerebellar ataxia. CA= Cayman ataxia. MSS= Marinesco-Sjögren syndrome (from Fogel and Perlman, 2007).

Despite the higher than expected genetic heterogenicity of SCARs, the involved molecular pathways can be clustered in three groups (Figure 2): i) Mitochondrial metabolism (e.g. *FXN*, *SACS*, *SPG7*, *COQ8A*) ii) DNA repair/genome stability (e.g. *ATM*, *SETX*, *APTX*, *MRE11A*) iii) Lipid metabolism (e.g. *NPC1*, *CYP27A1*, *PNPLA6*, *PLA2G6*) (Synofzik et al., 2019). Therapies are available for SCARs due to enzymatic deficiency that can be treated by supply of lacking products, for example α-tocopherol for Ataxia with isolated vitamin E deficiency (MIM# 277460) caused by mutations in *TTPA* (Gabsi et al., 2001), and CoQ_{10} for Primary coenzyme Q deficiency due to *COQ8A* mutations (MIM# 612016) (Mignot et al., 2013), but most SCARs still lack therapies. However, as well as for SCAs, treatments are under development, including modification of disease pathways, restoration of disease proteins, and gene therapy approaches (Synofzik et al., 2019).

1.2 Next Generation Sequencing (NGS)

1.2.1 From Sanger sequencing to massive parallel sequencing: overview of NGS

Frederick Sanger and collaborators developed in 1977 (Sanger et al., 1977) a method to sequence DNA, a milestone in genetics and still widely used. Sanger method is also called "chain termination method", since it is based on the principle that the random incorporation of dideoxynucleotides cause the termination of the DNA chain elongation, generating DNA fragments of various sizes that can be therefore separated through gel (at the beginning) or automatic capillary (later) electrophoresis. Sanger sequencing is a very precise and high-quality method to read DNA sequences up to 1-2 kb, and is still widely used for many purposes, but it almost so far obsolete for application in medical genetics, since it is very expensive and time consuming. Indeed, until roundabout 2013 the routinely screening of causative gene of genetic disorders was made gene by gene, leading to a very challenging research of candidate genes for patients' diseases. Moreover, it was prohibitive to apply Sanger sequencing in large-scale screening of human DNA such as the Human Genome Project which took fifteen years to be completed having to rely on Sanger sequencing applications in hundreds of laboratory teams worldwide.

In 2005, the first NGS platform (454 Life Science) was developed (Margulies et al., 2005) (Morganti et al, 2019), making DNA sequencing time- and cost-effective. Later, additional companies embark developing more precise and cost-effective NGS platforms. Despite only fifteen years has passed since the development of the first NGS sequencer, third- and fourth-generation sequencers are being developed with the ability to produce huge amount (hundred or even thousands of billions gigabases) of data. Unexpectedly, the technology has by far overtook our

capability to understand what it produces, therefore big data analysis and interpretation is "the challenge" of current and future NGS era (Morganti et al, 2019).

NGS procedure can be summarized in three steps, independently by the platform considered (Figure 3):

- Library preparation: libraries are created by fragmentation of DNA samples. Since lot of samples are often sequenced together, fragments of each library are linked to adaptors in order to make every fragment of each library recognizable. Fragments with their adapter are then amplified depending on the platform used for sequencing. Amplification is almost always present during library preparation, due to necessity to have every fragment reasonably represented in libraries, but also library preparation so-called "PCR-free" exists, principally for whole genome sequencing.

- Sequencing: it can be short read (100-500 bp) or long read (more than 500 bp). Short read is widely used specially in clinical research, due to its low cost and high accuracy, but it is not able to read repetitive regions, that can be sequenced by long-read method instead. Method used for sequencing is variable depending on the platform chosen (e.g. sequencing by synthesis, by ligation, pyrosequencing)

- Data analysis: NGS produces a huge amount of raw data. It is crucial for those who are going to interpret these data to filter all these informations. Typically, there are four steps in a basic data analysis workflow:

  - Base calling: this is the translation of signal acquired during sequencing in sequences of bases.

  - Reads alignment: sequences (called "reads" in NGS language) are aligned to a reference genome. Short-read sequencing usually produce hundreds of billions of sequences, so a single base can be read even thousands of times leading to very high accuracy.

  - Variant calling: differences between overall sequencing and reference genome are identified. Possible sequence variants are single nucleotide variations (SNVs), insertions/deletions (indels), and copy number variations (CNVs). Short-read sequencing have a very high accuracy in detecting SNVs and small indels but has big difficulties in identification of large repeated regions and CNVs, that can be detected often with the only help of powerful software. Long-read sequencing do not usually generate the same huge amount of reads as short-read does, so has lower

accuracy in detecting SNVs, but is more useful to detect indels and CNVs because of higher length of reads.

- o Variant annotation: this is the only step that is not automatized, since whoever do the analysis has to recognize variants with a potential biological value.

Correct filtering and proper interpretation of huge amount of data is probably the bigger issue that all users have to face. Improvements in big data analysis science and bioinformatics will be likely crucial to deal with this challenge.
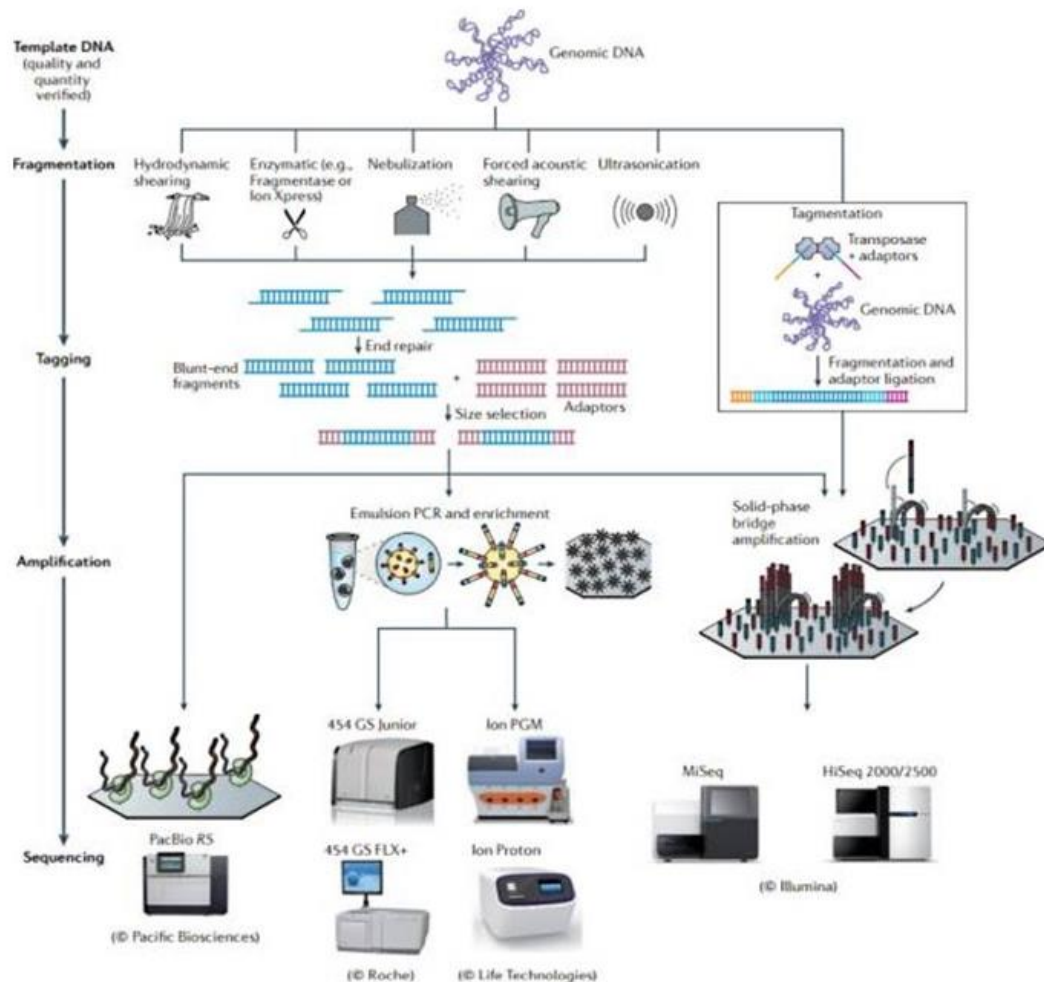
Figure 3. Overview of main NGS platforms available. Despite the different types of sequencing chemistries employed by each NGS sequencing platform, libraries preparations for massive parallel sequencing follow common major steps: fragmentation of template DNA, end repair and indexing, and amplification. (from Applications of Clinical Microbial Next-Generation Sequencing: Report on an American Academy of Microbiology Colloquium held in Washington, DC, in April 2015).

1.2.2 NGS applications

NGS wide-commercial applications were firstly concerned to overtake the gene-by-gene approach in the clinical setting. Advent of NGS in the detection of mutations in genes involved in inherited diseases brought a huge change in the approach of identifying new etiologies in Mendelian diseases in terms of costs and time. The first successful NGS application took to the identification of *DHODH* gene, cause of Miller syndrome (MIM# 263750) (Ng et al., 2010). Subsequently, several new disease-causing genes were identified, even in conditions where the traditional linkage analysis would have been a true challenge (Wang et al., 2010) (Isidor et al., 2011). Several NGS applications exist, but can be summarized in three main groups (Morganti et al., 2019).

- Genomics: as previously mentioned, NGS first goal was to overtake the gene-by-gene approach in the clinical field in order to find mutations in a time- and money-safe perspective, so the genomic applications were the first to be used and are also so far the most common and the best validated NGS methods. Main genomics methods are:
  - Target resequencing panels: only a custom number of genes are sequenced. This approach allows focusing resources and timing in genes of interest (e.g. all genes involved in a certain disorder) with a high coverage (200-1000 X coverage or more).
  - Whole exome sequencing (WES): all the coding part of the genome (1-2%) is sequenced. This method allows finding new correlations between genes and disorders, and also to broad the research of disorders etiology to a larger number of gene when all genes typically involved in a certain disorder have been already tested. Indeed, WES is sometimes used as a second-tier approach when patients result to be negative direct sequencing of few genes of interests or to multi-gene panels, but its coverage is significantly lower (30-50X) and sometimes certain regions are not sequenced with appreciable coverage (<20X) or even not sequenced.
  - Whole genome sequencing (WGS): the entire genome is sequenced, but coverage is usually very low (5-20X). Since everything here is sequenced, WGS is a "capture-free" method. Because of its purpose and output, WGS can be considered a "once in a lifetime" test.
- Transcriptomics: the transcriptome (RNA sequencing or RNA-seq) is the full set of transcripts in a single cell or in a population of cells at a specific time point and it is used to profile molecular signature and gene expression in specific conditions. Moreover, RNA-seq can help to detect at the same time SNVs, gene expressions, gene fusions, and more (Wang

et al., 2009). This method has become also popular for discovery and profiling of non-coding RNA (ncRNA) and microRNA (miRNA).

- Epigenomics: epigenome is the full set of inheritable features not caused by changes in DNA sequence. Most common epigenetic traits are DNA acetylation, DNA methylation and histone-modification. Complete methylation profile (also called methylome) and Chromatin Immunoprecipitation Sequencing (ChIP-seq) are the most commonly used NGS applications in epigenomics. Epigenetics is nevertheless important for the discovery of new mechanisms underlying genetic disorders.

1.2.3 The role of NGS in HAs

Since its advent, NGS became a powerful tool for large screening in patients with genetic disorders. Screenings in little, mid and large cohorts of patients were very helpful to understand which applications better suits to achieve the higher diagnostic yield, and to re-write the epidemiology of inherited rare diseases that were sometimes considered confined to specific populations (Ruano et al., 2014). Nonetheless, application of NGS in HA had a huge impact. First, it took to an exponential discovery of new HA genes in a short time, contrary to the age of positional cloning (Orr et al., 1993), and Sanger sequencing (Galatolo et al., 2018) (Figure 4).
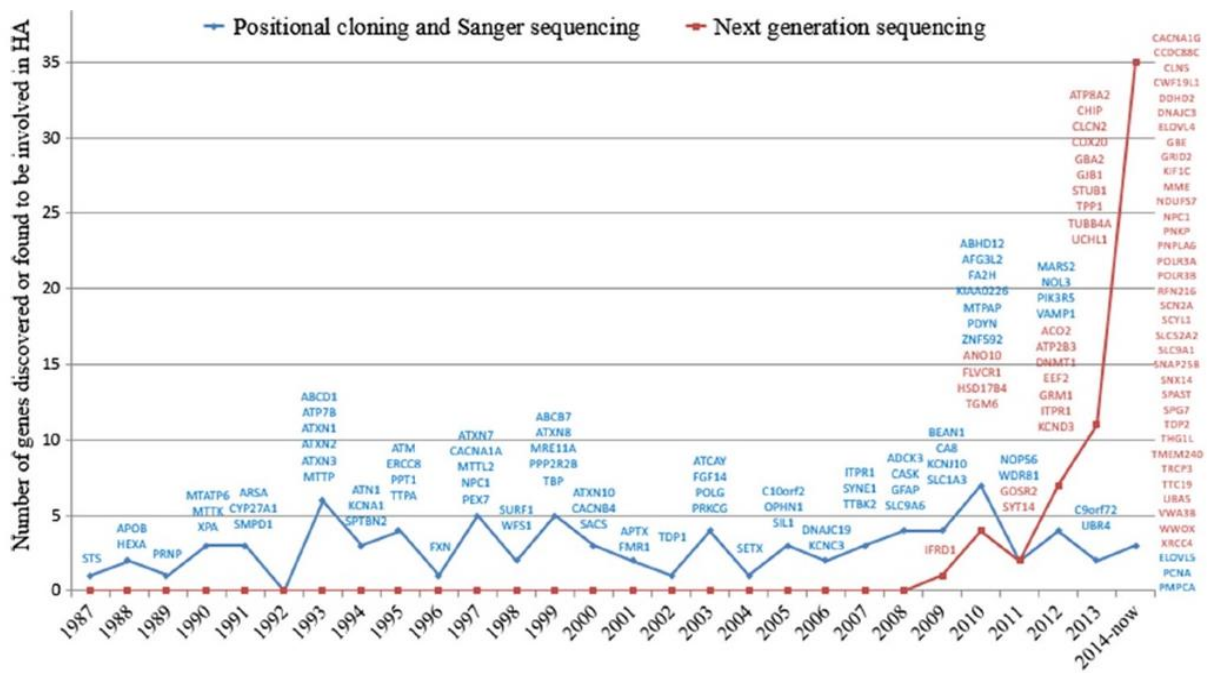
Figure 4. Timeline of the discovery of genes involved in hereditary spinocerebellar ataxias. Comparison between hereditary ataxia (HA) genes discovered by positional cloning/Sanger sequencing (indicated in blue) and HA genes identified through NGS (red) shows that massive parallel sequencing had a pivotal role in the exponential increase of genetic etiology in HA since early 2010s (from Galatolo et al., 2018).

Second, it permitted to expand the genetic heterogeneity of several neurological disorders, including HAs. As an example, NGS permitted to define *SPG7*, originally believed to account only for 5-12% of recessive HSPs, as one of the most common cause of pure cerebellar ataxia (van de Warrenburg et al., 2016). Besides, NGS could also recognize a new role in other disorders for HAs genes (e.g. *SYNE1*, currently a common cause of spinocerebellar ataxia, but recently also with a role in corticospinal tract degeneration) (Synofzik et al., 2016).

Nucleotide repeat expansion disorders are the most common cause of HAs (Durr, 2010), even if account for a limited amount (2-13%) of late-onset sporadic cases (Abele et al., 2002), and GAA intronic expansion associated to FA are the most common by far in recessive HAs. This actually represent a critical issue in molecular diagnosis, considering that common NGS strategies used in clinical scenarios (TRP and WES) cannot detect repeated expansions. Thus, it is recommended to test at least the most common loci associated with nucleotide expansion (SCA1, SCA2, SCA3, SCA6, SCA7, SCA8, SCA10, SCA17, DRPLA and FA) before application of any NGS methods.

Recently, we reviewed the state of the art in gene testing in HAs (Galatolo et al., 2018) reviewing data from 12 NGS study performed in HAs cohorts. Our study concluded that exome sequencing as first-tier approach has a higher diagnostic yield compared to multi-gene panels (on average 36% vs 17% positive results) and should be preferred in clinical setting. We also observed that more than a half of disease-causing mutations are localized in a restricted pool of genes and those should be scrutinized first while filtering data in exome sequencing. Our data well correlate with more recent studies of the exome in large cohort of HA patients (Coutelier et al., 2018) (Sun et al., 2019).

1.3 Models *in vitro* and *in vivo*

1.3.1 *Saccharomyces cerevisiae* as a tool to predict *in vivo* pathogenicity of human mutations

Saccharomyces cerevisiae, commonly known as baker's yeast, is a unicellular eukaryote considered for many years a milestone in biologic research since its small genome (about 12.000 kb and 6000 genes) and several practical issues (Duina et al., 2014). Yeast DNA was used at the dawn of recombinant DNA era (Curran et al., 2014) and the genome of *S. cerevisiae* was the first to be sequenced among eukaryotes (Goffeau et al., 1996).

Besides both astonishingly functional and amino acid sequence conservation between *S. cerevisiae* and higher eukaryotes, low costs and ease to maintain made yeast a powerful *in vivo* platform for fast and efficient determination of pathogenicity of human mutations (Duina et al, 2014). This is made possible by using complementation assay, a technique which assess the capability of yeast cells to survive and grow only when the selected mutant allele (carrying the human gene harboring the mutation of interest) is present (Figure 5). Since the gene of interest could be necessary for survival, a wild-type copy of the studied gene (see in figure 5, green) is expressed from a maintenance vector, that also encode an enzyme (*URA3*, blue in figure 5) lethal for yeast in presence of 5-fluoroorotic acid (5-FOA), in haploid yeast disrupted for the gene of interest. The allele selected, harboring human mutation of interest (red in figure 5), is then expressed with a second vector carrying also a *LEU2* gene (yellow in figure 5) used to select yeast in medium lacking uracil and leucine. Successive growth on medium with 5-FOA permit selection of yeast cells that lost the maintenance vector so that growth effects are only caused by the allele carrying the human mutation (Oprescu et al., 2017). This relatively simple assay allows to screen quickly for pathogenicity of human mutations, provided that the gene of interest is expressed in baker's yeast.

 Despite obvious limitations of the assay and the necessary caution in considering yeast as surrogate of differentiated cells in a multi-cellular environment, yeast remains a powerful high-throughput model to predict *in vivo* the effect of human mutations in some forms of HA.
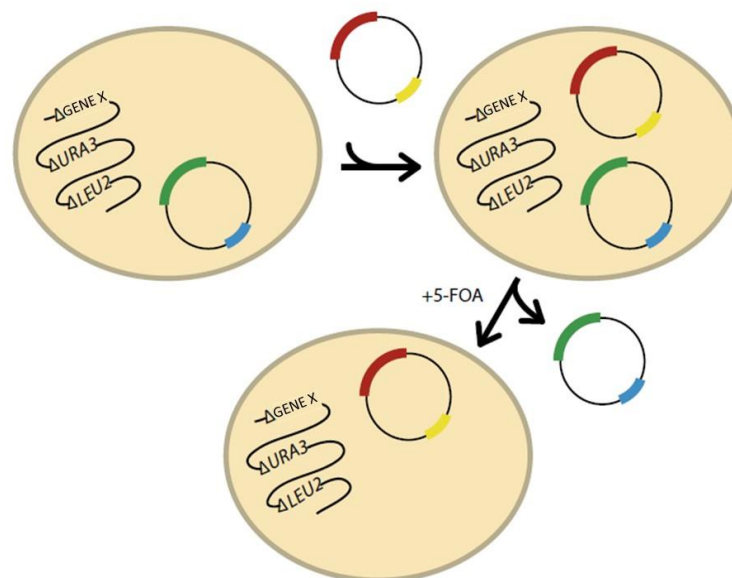
Figure 5. Overview of the yeast complementation assay. Haploid yeast for endogenous gene of interest (Gene X) is maintained viable with a wild type copy of that gene (green) and substituted with mutant alleles (red). A system involving *URA3* (blue) and *LEU2* (yellow) genes is used for selection of colonies first in medium lacking uracil and leucine, and than in one containing 5-fluoroorotic acid (5-FOA). All genes mentioned are included in plasmid constructs (see text for detailed mechanism description). (modified from Oprescu et al., 2017).

1.3.2 *Drosophila melanogaster* as a model to study neurodegenerative disorders

Drosophila melanogaster are little flies that belong to Diptera order and Drosophilidae family. Used in the biological research since the early XX century (Morgan, 1910), Drosophila represents a powerful model system in biology since it is easy and cheap to maintain, its genome is compact and easy to manipulate, is simple to handle, and remarkably, the level of similarities with human genome is high with many orthologous related to human neurological disorders.

One of the greatest advantages of Drosophila is the short life cycle that leads to a large progeny in a short time. Development to adult from a fertilized egg takes about 10-12 days at 25°C, but temperature influences the speed of this processes: at 29°C adults emerge from pupal cases in about 7 days, and in 20 when flies are raised at 18°C. Embryogenesis takes 1 day, other 3 days are required to complete three larval stages (first, second and third instar larva) before metamorphose in pupa for 5-6 days and subsequentially eclosion (Hales et al., 2015), after which males are already sexually mature, whereas females are virgins for about 8 hours (Figure 6). The fly genome is highly compact and organized in 4 pairs of chromosomes with a size of about 180 megabases that comprises about 14.000 genes with low redundancy (Bosco et al. 2007). Moreover, 2 out of 3 human disease-causing genes are estimated to be represented in the genome of *D. melanogaster* (Reiter et al., 2001).

Drosophila contributes heavily to the study of human neurodegenerative disorders because fruit flies share basic functional, developmental, of molecular neurobiological principals with humans (McGurk et al., 2015). Furthermore, it possesses a complex central nervous system with neurons, glia, blood-brain barrier, and structures that are increasingly associated with other mammalian brain structure (e.g., mushroom bodies, recently considered the corresponding of cerebellum; Farris 2011) and it can perform complex behaviors such as circadian rhythm, learning, memory, aggressivity and sleep, although the measurements of these behaviors is not easy (Prüßing et al., 2013). Therefore, Drosophila has represented an attractive model system for conditions such as Alzheimer's disease, amyotrophic lateral sclerosis (ALS), Parkinson's disease, polyglutamine related diseases (including most common SCAs), and several motor neuron diseases (McGurk et al., 2015). Furthermore, assays in fruit fly are available to study all tissues and tools of investigations are relatively simpler than manipulating human cells or rodents (Ugur et al., 2016). For example, studies of neuronal loss and degeneration, brain vacuolization, muscle integrity, retinal degeneration, survival and locomotion analysis are easy and common studying fruit fly research (see McGurk et al., 2015 for all references) (Figure 6). Hence, with a nearly uncountable number of lines, tools and techniques available to study this model, and a very large online community to share materials and support

innovations, it is understandable that *Drosophila melanogaster* represents one of the most powerful models in biology.
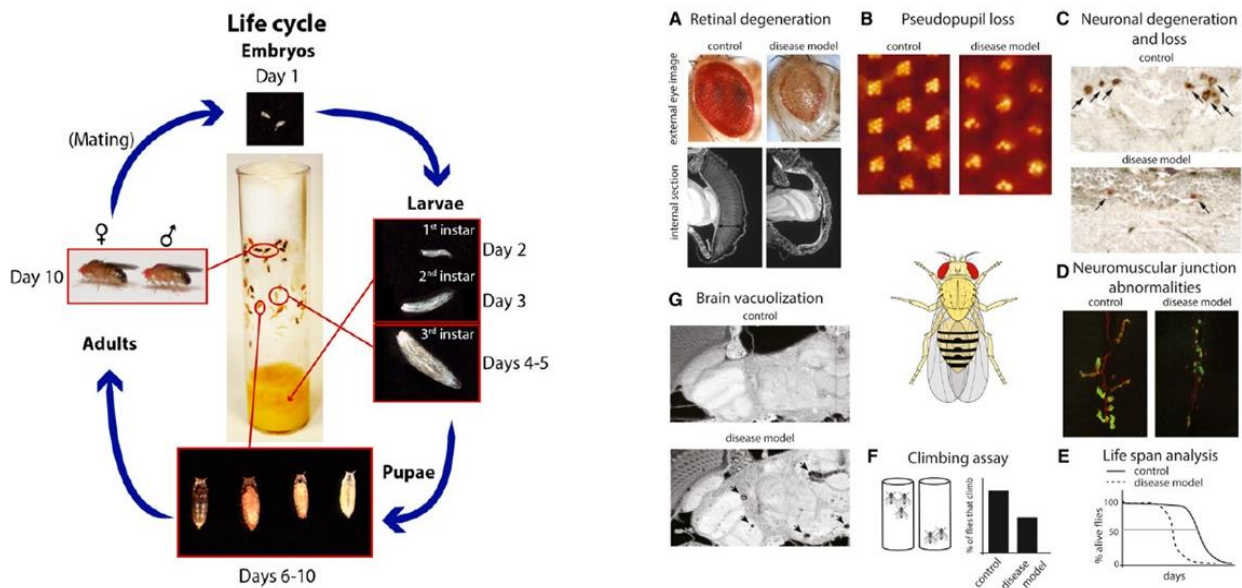
Figure 6. Life cycle and examples of robust assays to assess neural degeneration and dysfunction in Drosophila melanogaster. Life cycle in Drosophila (left picture). Fruit flies are grown in vials with food at the bottom. At 25°C, flies spend around 9-10 days in larval and pupae stages before maturation in adult and eclosion from pupal cases. First two larval stages (called 1st and 2nd instar) are spent in the food. 3rd instar larvae crawl out of the food to pupate on the side of the vial. During pupae stages metamorphosis occur, and the darkening wings within the pupal case indicate that maturation is nearly complete. Examples of robust assays to assess neural degeneration and dysfunction in Drosophila (right picture). (A) External eye (upper panels) offers a quick readout of retinal degeneration, whereas internal sections (lower panels) allow examination of the retinal tissue. (B) The pseudopupil assay permits a quantitative measure of degeneration in adult flies by assessing the structure of the photoreceptor cells by counting the number of intact rhabdomeres (the specialized organelles of the photoreceptor cells). (C) Neuronal degeneration and loss can be assessed by histochemistry and immunohistochemistry. (D) Immunostaining for presynaptic and postsynaptic structures enables to detect neuromuscular junction abnormalities. (E) Integrity of fruit flies can be assayed checking their lifespan. (F) Neuromuscular defects can be determined testing locomotion by negative geotaxis assay, that measures the ability of flies to climb. (G) Brain degeneration can be estimated by formation of vacuoles in brain sections. (left figure from Hales et al., 2015; right figure from McGurk et al., 2015).

1.3.3 *Danio rerio* as a model to assay neuromuscular functions

Danio rerio, commonly named as zebrafish because of its typical striae, is a small teleost fish (less than 5 cm in adulthood) widely used in biological research in recent years and originally proposed as model of human conditions in early 1960s (Grunwald and Eisen, 2002). Besides easy handling, small size, low costs and easy maintenance and breeding, two major peculiarities made this animal worth for human research: i) it is a vertebrate but also has some practical characteristics typical of

invertebrate models (Langherinrich, 2003) and ii) its embryos are translucent, allowing the monitoring of the whole organisms during the development (Rico et al., 2011) (Parker et al, 2012) (Stewart et al., 2014) (Stewart et al., 2015) (Fontana et al., 2017). Moreover, zebrafish embryos can be easily grown in 96-microplates wells, making this model suitable for easy pharmacological and genetic screening by just adding few drops of drugs or chemical in sea water (Rosemberg et al., 2012). This approach is particularly attractive since it offers the opportunity of wide (even high-throughput) chemical genetic screen *in vivo* using living embryos without prior knowledge about anything regarding molecules tested.

Another advantage of zebrafish is its high degree of genomic sharing with mammals (MacRae and Peterson, 2015). Genome is conserved and about 70% of zebrafish genes have at least one human orthologue, and nearly half of human genes have a one-to-one relationship with zebrafish orthologue (Howe et al., 2013). About 26.000 protein coding genes are counted, presenting a significative percentage of duplicate genes. Genome redundancy can be useful in evolutionary studies, but it can make harder studies of molecular pathways or investigation in gene-related disorders since could potentially bypass induced DNA damage or silencing (Fontana et al., 2017). Embryogenesis occurs at 26-28°C, and embryonic development is very fast, reaching the plan of body development 24 hours-post-fertilization (hpf). Neurogenesis begins at 10 hpf, synaptogenesis and first behaviors at 18 hpf, whereas hatching occurs at 52 hpf (Kabashi et al., 2011) (Figure 7A).

Zebrafish is a powerful animal model not only for the easy to study of neuromuscular system (Langheinrich, 2003) (Figure 7B), but also for its use as model of social, affective and motor behaviors in a wide range brain disorders (Egan et al., 2009) (Kalueff et al., 2013) (Norton, 2013) (Blaser and Vira., 2014). Moreover, easy, quick and validated tests are available in embryos to assay neuromuscular function, which is very helpful for fast screening of potential genes involved at any point of motor unit pathways. To date, zebrafish is successfully used as model for autism spectrum disorders, intellectual disabilities, schizophrenia, cerebellar ataxias, hereditary spastic paraplegias, Charcot-Marie-Tooth disease (see Kozol et al., 2016 for references), epilepsy, Alzheimer's disease, Parkinson's disease and affective disorders and more brain diseases (see Fontana et al., 2017 for references).
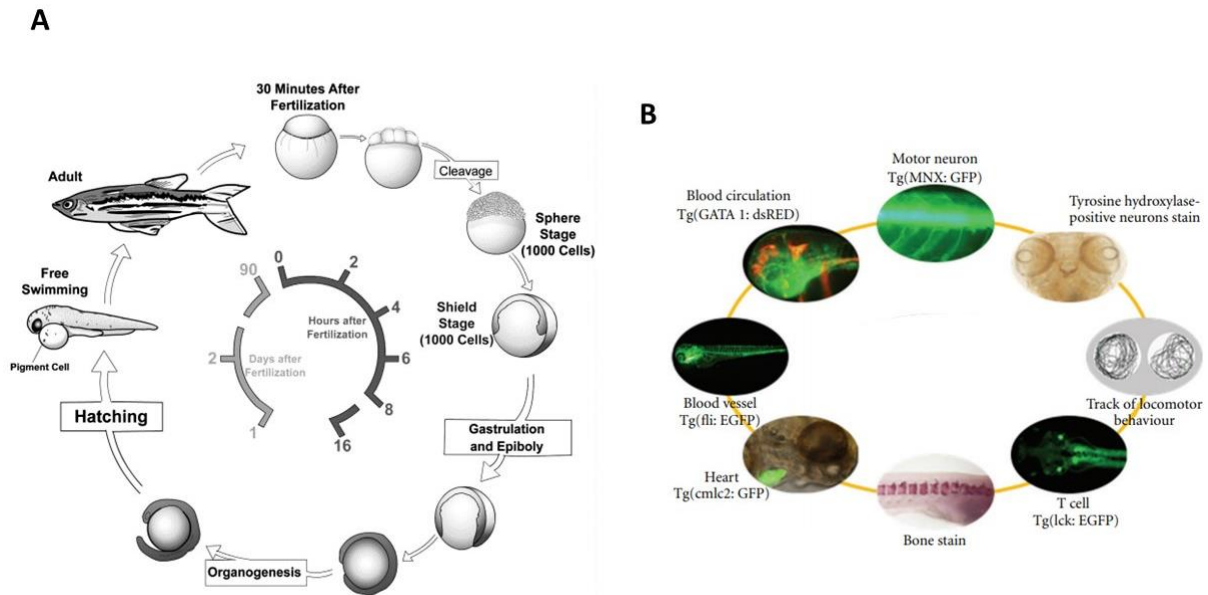
Figure 7. Life cycle, structures and features available for studies in Danio rerio. (A). Life cycle of *Danio rerio*. Zebrafish develop rapidly from a one-cell zygote that sits on top of a large yolk cell. Gastrulation begins approximately 6 hours post fertilization, hatching at 2 days as a free-swimming larvae. Zebrafish reach sexual maturity around 3 months of age and can live for up to 5 years. (B) Example of tissues and assays that can be used in zebrafish. Readouts on live larvae consist in the assessment of zebrafish integrity by checking their life span, and in determination of locomotor defect measuring coiling frequency, touch-evoked response, and movements tracking. Many tissues and districts can be studied in zebrafish, as indicated in the figure. (part A from D'Costa and Sheperd, 2009; part B modified from Hung et al., 2012).

1.3.4 Skin fibroblasts as an *in vitro* model of patients' mutations

Primary skin fibroblasts represent an essential system to study human disorders since represent the *in vitro* transposition of patients' genome and likely their phenotype. Easily obtained with diagnostic punch biopsies, establishing a permanent cell lines from the patient harboring a rare inherited disorder is invaluable and offers new opportunity to investigate mechanisms as well as the consequences of gene mutations at the mRNA and protein levels (Auburger et al., 2012). Furthermore, over the aforementioned advantages, fibroblasts are quite easy to genetically manipulate and hence prone to gene editing and silencing, but optimal use of them require caution due to their aging passage by passage, or to suboptimal matching of patient cells with control cells and other conditions that may generate hardly reproducible results.

More importantly, patient's derived skin fibroblasts represent also a source for generation of induced pluripotent stem cell (iPSC) that can be afterwards differentiated in cells of the tissue of interest (e.g. neurons for studies of neurodegenerative diseases) (Takahashi et al., 2007), although the generation of isogenic controls would be ideally needed for precise investigations.

1.4 Gene silencing and gene editing

Gene silencing is a generic term to describe the repression of gene expression. Widely known mechanisms are antisense oligonucleotides (ASO) and RNA interference (RNAi), which may act on either microRNAs (miRNA) or double stranded RNA (dsRNA) to downregulate or abate gene functions. miRNAs are short endogenous non-coding RNA normally active in gene expression regulation, whereas dsRNA are synthetic molecules used for similar purposes (Heidenreich and Zhang, 2015).

ASO are delivered as single strand and find their target alone, so they can make steric blocks, modulate splicing, and recruiting protein factors (Watts and Corey, 2012). Gene editing is a generic term to describe permanent modification in the genome. Widely known genome-editing tools are zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs), both recognizing DNA sequences through protein-DNA interactions, and CRISPR-associated (Cas) effector proteins of clustered regularly interspaced short palindromic repeat (CRISPR) systems, that requires one or two guide RNAs to target a specific locus.

1.4.1 The GAL4-UAS system

The UAS-GAL4 system is a method used to study gene expression and was developed in early 1990s (Brand and Perrimon, 1993). This bipartite system consists of two parts: the *GAL4* gene that encodes a yeast transcription activator, and the gene or a construct of interest downstream of the *UAS* promoter (Upstream Activation Sequence), a short sequence that specifically binds GAL4 to activate gene transcription. Even if available for several models, GAL4-UAS system is typically employed in fruit fly research (Caygill and Brand, 2016). GAL4-UAS system is hence used to activate the expression in a tissue-specific manner. Indeed, thousands of Drosophila transgenic GAL4 lines, called "driver line", have so far been created, each expressing GAL4 only in a certain tissue or subset of cells. Transgenic "driver lines" (containing *GAL4*) are then crossed to a second transgenic line carrying a UAS-dependent transgene (hence a gene downstream the *UAS* sequence), termed "responder line", so that the progeny will express the UAS-dependent transgene only in tissue/cells where GAL4 is also expressed (Figure 8A).

GAL4 "driver lines" and UAS "responder lines" are generated and maintained as separate stocks, so GAL4-UAS system cannot be active in these stocks, making these flies completely normal. Moreover, the presence of *GAL4* and *UAS* itself in driver and responder lines has no effect. This system can thence lead to tissue/cell specific expression of UAS-dependent transgene. Applications

may vary depending on the gene downstream to *UAS* sequence. If an endogenous gene is downstream *UAS* sequence, an overexpression of that gene will occur, whilst an ectopic expression takes place if a non-endogenous is inserted. Nevertheless, genes reporter like GFP or protein coding for synthesis of toxic or pharmaceutical compounds for high-throughput screening applications can be used as well. On the other hand, the most common application is the use of RNAi to silence genes of interest and study tissue/cell specific loss-of-function. Indeed, with the ever expanding Transgenic RNAi Project (TRiP) tools an increasing number of transgenic RNAi fly stocks for almost all genes are available to the community. All this reflects the powerfulness of the tool to study the role of genes and their involvement in biological pathways or in human disorders as well.
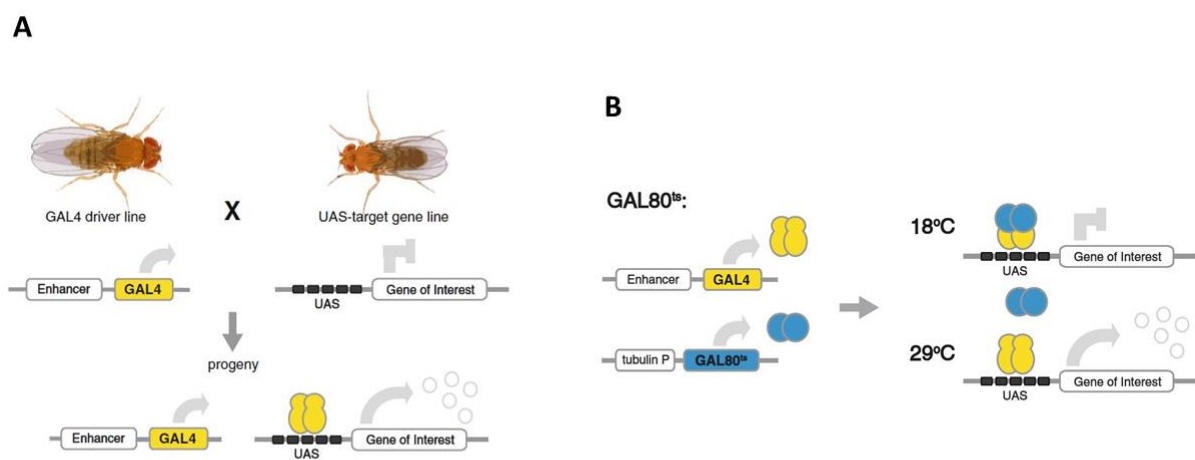
Figure 8. The GAL4-UAS system. (A) The basic GAL4-UAS system. GAL4 driver and UAS-target gene fly lines are generated and maintained as separate stocks. In the absence of GAL4 there is no expression of the target gene. Crossing a fly expressing GAL4 to a fly carrying a UAS-target gene results in targeted gene expression in the progeny of the cross. (B) Modification to the basic GAL4-UAS system. The presence of a ubiquitously expressed temperature-sensitive allele of GAL80 (*tubGAL80^{ts}*) will result in inhibition of GAL4 activity at the permissive temperature of 18°C. At the restrictive temperature of 29°C GAL80^{ts} cannot bind GAL4, therefore GAL4 activity is not inhibited and GAL4-dependent transgenes will be expressed (part A and B from Caygill and Brand, 2016).

Despite the aforementioned basic function and employment of GAL4-UAS system, several useful variations of this system are also available (Caygill and Brand, 2016). For instance, the use of GAL80^{ts} (ubiquitously expressed under control of the tubulin 1α promoter) transgenic lines can be used to limit the activation of GAL4-UAS system in selected time windows. Indeed, GAL80 protein can repress the interaction between GAL4 and *UAS* by binding GAL4 protein, so that GAL4 no longer binds the *UAS* sequence (Nogi et al., 1984) (Suster et al., 2004). Combining GAL4 and

GAL80[ts] transgenic lines, expression of GAL4-UAS system can be modulated. This variation takes advantage of a temperature sensitive form of GAL80 (*GAL80[ts]*) related to the p.Gly203Arg mutation in the GAL80 protein (Matsumoto et al., 1978). GAL80[ts] cannot bind GAL4 above 29°C, but preserves its repressive function at 18°C (Figure 8B). Activation or repression of GAL4-UAS system can hence be easily achieved just shifting temperature of flies' environment. These and much more applications and variations of GAL4-UAS system are available making this approach a masterpiece for Drosophila research, and encouraged also translating this system to other model organisms (Caygill and Brand, 2016).

1.4.2 The CRISPR/Cas9 system

The Clustered Regularly Interspaced Palindromic Repeats (CRISPR)/CRISPR-associated protein (Cas) system is a mechanism so far worldwide used in genomic engineering "borrowed" from prokaryotes that use it as innate immunity system, to protect themselves by viral DNA infections (Barrangou et al., 2007). Indeed, once the infection occurs, bacterial nucleases fragment viral DNA and incorporate it as short sequences (about 20 bp length) separated by short palindromic repeat, that will be used to recognize future invasions of the same type of "invader".

Despite this bacterial mechanisms require several components, researchers achieved to an engineered application using only two components (Jinek et al., 2012) (Mali et al., 2013), that are the nuclease Cas9, an enzyme that cuts DNA, and guide RNA (gRNA), a RNA molecule (20 bp) that interacts with Cas9 and guide it to the exact point to cut. gRNA can guide Cas9 to a precise point since its sequence is homologous to a specific target DNA sequence and immediately upstream of a three base pair DNA protospacer adjacent motif (PAM) —whose sequence for *Streptococcus pyogenes* Cas9 (the most commonly used) is NGG—, a sequence necessary to Cas9 to cut DNA (Williams and Warman, 2017). Once Cas9 cuts DNA through a double-strand break, physiological DNA repair mechanisms occur. If DNA is repaired by non-homologous end joining (NHEJ) breaks are often repaired after filling the gap with variable numbers of base pairs, so that insertion and/or deletion in frame or frameshift can be created (Figure 9A, left) If homology-directed repair (HDR) occurs instead, a template for the region with the breakpoint is used to repair the damage, but no errors occur in this case (Williams and Warman, 2017) (Figure 9A, right).

In normal conditions, cells choose HDR in case of single strand breakage, since the remaining strand can be used as template, whereas NHEJ is used when there is a double strand breakage, since there is no a wild type template available in that case. In genetic engineering, NHEJ is induced

through a double strand breakage caused by the Cas9 guided by one gRNA, if complete disruption (knock-out, KO) of a gene wants to be achieved. In this case, since indel entity is unpredictable, is necessary to genotype all the progeny in order to distinguish strains harboring early frameshift mutations leading to premature truncated protein. On the other hand, HDR can be used to insert a certain DNA sequence with two double strand breakage created by Cas9 guided by two gRNA, and DNA sequence of interest that can be inserted through homology of sequences flanking both breakage and insert (contained in a vector). In this way, an entire sequence (e.g. a non-endogenous gene, a reporter gene, a sequence containing a human mutation leading to possible creation of knock-in models) can be inserted without causing unpredictable errors (Ran et al., 2013).
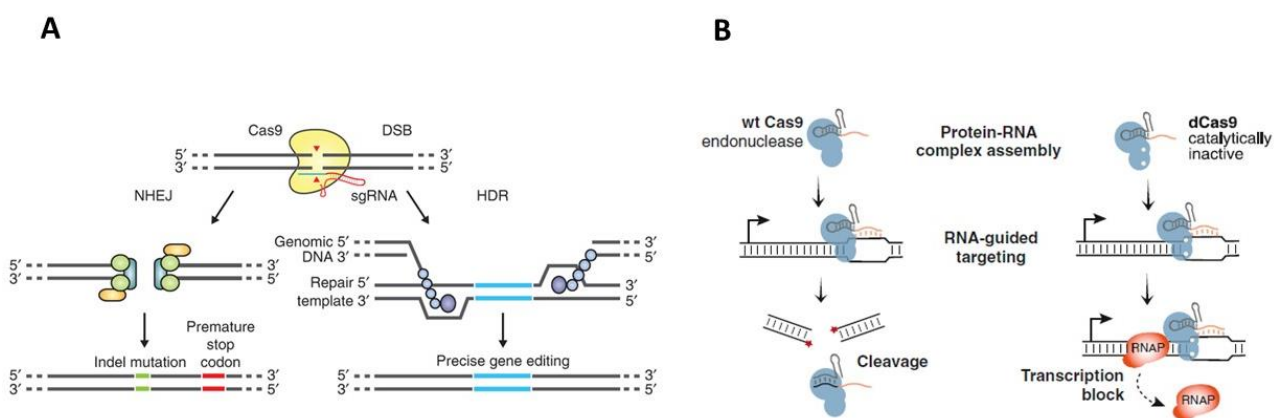
Figure 9. The CRISPR/Cas9 technology. (A) Double strand breakage (DSB) repair promotes gene editing. DSBs induced by Cas9 (yellow) can be repaired in one of two ways. In the error-prone non-homologous end joining (NHEJ) pathway, the ends of a DSB are processed by endogenous DNA repair machinery and rejoined, which can result in random indel mutations at the site of junction. Indel mutations occurring within the coding region of a gene can result in frameshifts and the creation of a premature stop codon, resulting in gene knockout. Alternatively, a repair template in the form of a plasmid or single-stranded donor oligonucleotides (ssODN) can be supplied to induce the homology directed repair (HDR) pathway, which allows high fidelity and precise editing. Single-stranded nicks to the DNA can also induce HDR. (B) Comparison between mechanisms of wild type and dead Cas9 (dCas9). Wild type Cas9-gRNA complex (left) binds and cuts specific DNA targets due to the nuclease activity of the Cas9 protein. dCas9 (right) is still able to form a complex with the gRNA and to bind to specific DNA target, but when the targeting occurs on the protein-coding region, it blocks RNA polymerase (RNAP) and transcript elongation due to its loss of cleavage action. White dots in the dCas9 indicates the mutations that abolish Cas9 nuclease activity (part A is from Ran et al., 2013; part B is from Qi et al., 2013).

CRISPR/Cas9 system is such a revolutionary tool that has been deeply studied in recent years, and several applications and variants are so far available. For instance, dead Cas9 (d-Cas9) is a Cas9 variant harboring two missense mutations in residues very important in both Cas9 endonuclease domains, so that d-Cas9 loses its enzymatic activity, but still can bind DNA thus altering gene

expression by steric hindrance since denies proteins like transcription factors and polymerases to access DNA (Qi et al, 2013).

CRISPR/Cas9 system can be microinjected into fertilized eggs to produce genetically modified offspring of model animals easily and quickly (Williams and Warman, 2017).

On the other hand, potential off-target effects are the real limitations of this technology. Indeed, is it possible that Cas9, similarly to other nucleases, could cut in unexpected DNA regions. Performing complete genome sequencing afterwards every gene editing event would be ideally the only way to verify with certainty the absence of off-targets, but this approach is clearly impracticable. However, current CRISPR/Cas9 *in silico* tools used to design gRNAs and to find correct cleavage sites, take advantage of several powerful approaches and algorithms to efficiently avoid off-target effects (Ran et al., 2013).

CRISPR/Cas9 system is so far a landmark in genetic engineering, and it is considered the milestone of so-called "precision medicine" (e.g. creation of knock-in models harboring specific human mutations) and future clinical applications in gene therapy.

2. Aims

 Aim of this project is to use NGS strategies to define the molecular characterization of 259 HA patients recruited mainly in three neurological collaborating centers. In this study we propose to use massive deep sequencing to screen patients with a clinical diagnosis of hereditary ataxia but negative to most common repeat nucleotide disorders, or refractory to previous gene-by-gene approach. To achieve this objective, we plan to use custom-designed multi-gene panels as a first-tier approach and exome sequencing as a second-tier approach in a restricted selection of cases refractory also to multi-gene panel application.

Three main purposes are at the base of this PhD project.

 The first one is to investigate what is the impact of NGS in the clinical setting of HA, trying also to derive whether peculiar features or phenotypic characteristics or type of inheritance allow to achieve a higher diagnostic rate and to stop HA patient's diagnostic odyssey.

 The second is to present the first NGS epidemiologic data of a well-characterized HA cohort in Italy, contributing to the global data provided by similar studies performed elsewhere in Europe and U.S. and contributing to the expanding information on early-onset ataxias gathered in the multicenter European project *PREPARE* (Preparing for therapies in autosomal recessive ataxias).

 The third goal is to use NGS as an output of mutations involved in new genes, or mutations in known genes associated with novel or atypical phenotypes. In this last goal, gene identification will be accompanied by functional research studies *in vitro* and *in vivo* models. Generation of disease models will also serve for potential future research of disease biomarkers or drug screening in HA. Through all these considerations, we will demonstrate that NGS is a crucial hub between the diagnostic and research faces of HA.

3. Materials and methods

3.1 Genetic analyses in HA patients

3.1.1 Sample collection and DNA extraction

Samples were collected mostly in three Italian centers (IRCCS Stella Maris Foundation, Pisa; University Federico II, Napoli; University La Sapienza, Latina) from patients with a clinical diagnosis of hereditary ataxia. Additional centers in Italy (all participants to the ITASPAX network, coordinated by Filippo M. Santorelli) contributed further samples for this study. All the participants provided written informed consent according to Italian National Health System guidelines and to Declaration of Helsinki. Genomic DNA was obtained from peripheral blood using MagPurix Blood DNA Extraction Kit 200 designed for the MagPurix DNA Extract (Zinexts, Zhonghe, Taiwan).

3.1.2 Preliminary analyses of repeated nucleotide expansions

Before undergoing massive parallel sequencing, all patients were tested for pathological trinucleotide expansions in SCA1, 2, 3, 6, 7, 8, 12, 17 using a TP-PCR based method (Warner et al., 1996) (Cagnoli et al., 2006), and for the intronic GAA expansion in *FXN* through a long-PCR established technique (Campuzano et al., 1996). Capillary sequencing for TP-PCR products were performed using 3130xl Genetic Analyzer (Thermo Fisher Scientific, Waltham, MA), and fragment analyses were performed with GeneMapper ID Software version 3.1 (Thermo Fisher Scientific).

3.1.3 Massive parallel sequencing

Four different custom multi-gene panels were employed in this study, since the discovery of new HA genes made necessary a periodic update of panel employed (full lists of genes are available in supplemental material).

- Ataxome 1.0, Haloplex technology (Agilent Technologies, Santa Clara, CA.), 82 genes.
- Ataxome 2.0, NimbleGen technology (Roche, Basel, Switzerland), 194 genes.
- Ataxome 3.0, SureSelect technology (Agilent Technologies), 273 genes.
- Ataxome 4.0, SureSelect technology (Agilent Technologies), 285 genes.

Two different commercially available exome sequencing approaches were also employed:

- Custom Constitutional Panel 17Mb (CCP17), SureSelect (Agilent Technologies), 2742 genes (Clinical exome).
- IDT, Nextera (Illumina, San Diego, CA), whole exome.

Custom panels from Agilent Technologies were designed using SureDesign (Agilent Technologies), whereas Roche panel was designed using NimbleDesign (Roche). In all designs, 50 bases upstream and downstream every coding exons were also covered, and designed probes were predicted to cover 99.5% of the whole region of interest in all cases. Preparations of libraries were performed following manufacture's instruction. In all systems, amplifications occur by PCR, custom probes are biotinylated, probe-target DNA hybrids are captured using streptavidin-coated beads, and sample-specific index sequences are incorporated in each library fragment, in order to make every sample recognizable after sequencing, since different samples are sequenced together. Indexes are usually added during a PCR step of the protocol, because index sequences are part of one or both primers. Main steps and differences are summarized below (Figure 10).

- HaloPlex Target Enrichment System (Agilent Technologies): 225 ng of genomic DNA (gDNA) per sample are digested using restriction enzymes, then gDNA restriction fragments libraries are hybridized for 16 hours to HaloPlex custom probes in a reaction that includes also indexes. Indeed, this is the only method, among the four here described, in which index sequences are included by ligation during hybridization reaction. Circularized probe-target DNA hybrids produced are then captured and amplified.
- SeqCap EZ HyperCap Library (Roche): 100 ng of gDNA per sample are enzymatically fragmented, then fragment ends are repaired and ligated to adapters. Fragments are subsequently amplified including indexes. Equal amounts of each indexed library are therefore mixed in one reaction, and 20 hours hybridization is performed using custom SeqCap custom probes. Hybridization products are thereafter captured and amplified once again.
- SureSelect Target Enrichment System (Agilent Technologies): 25 ng of gDNA per samples are enzymatically fragmented, then adaptors are added to ends of fragments. Adaptor-tagged gDNA libraries are amplified, and an 1-hour hybridization is performed using SureSelect custom probes. Libraries are afterwards captured and again amplified including index sequences.
- Nextera Rapid Capture Enrichment (Illumina): 100-500 ng of gDNA are tagmented, a process that fragments and tags the DNA with adapter in the same reaction using Bead-Linked Transposomes. Tagmented DNA is afterwards amplified including indexes, and

hybridization-capture step is performed twice (first hybridization 2-24 hours, the second 15-24 hours). After the second capture, libraries are amplified once more.
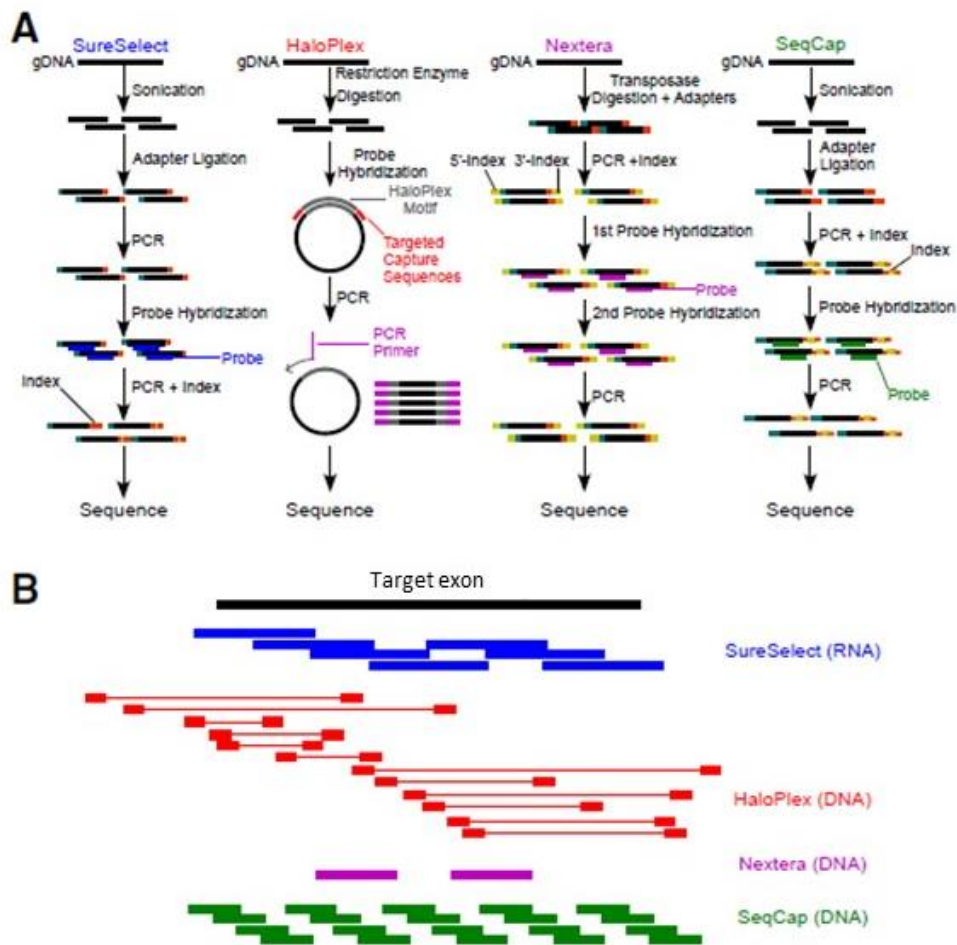
Figure 10. Summarized methods for targeted DNA sequencing. (A) Graphical depiction of four experimental protocols for library construction and capture between the four technologies. SureSelect and SeqCap have similar strategies for fragmentation and hybridization, whereas HaloPlex and Nextera use restriction enzyme and transposase to fragment DNA, respectively. (B) Distribution of probes relative to an example target exon. HaloPlex probes are only complimentary to sequences near the end of the probe, whereas the middle of the probe is a HaloPlex-specific motif (denoted by thick and thin red lines, respectively). SureSelect (blue) and SeqCap (green) use overlapping probes, whereas Nextera (magenta) uses gapped probes (modified from Samorodnitsky et al., 2015).

 Independently by the enrichment system used, amount of DNA adopted in each test was quantified using Qubit dsDNA BR Assay (Thermo Fisher Scientific). After pre-hybridization amplification, fragments lengths and libraries quantification were determined using Agilent TapeStation D1000 (Agilent Technologies), whereas Agilent TapeStation High Sensitivity D1000 (Agilent Technologies) was used after post-hybridization amplification. MiSeq (Illumina) was used to sequence libraries obtained from Ataxome 1.0 (Haloplex), Ataxome 2.0 (NimbleGen) and Ataxome

3.0 (SureSelect) using 6 pM, 8 pM and 9 pM of libraries, respectively, whilst NextSeq500 (Illumina) was employed to sequence Ataxome 3.0 (SureSelect) and Ataxome 4.0 (SureSelect) using 1.5 pM of libraries in each experiment. NextSeq500 was used to sequence the exomes.

3.1.4 Data analysis

Alignment to reference human sequence and variant calling were performed using Illumina Variant Call Analysis (Illumina) when Roche multigene panel or Illumina IDT exome were applied, whereas SureCall (Agilent Technologies) was used when Agilent panels or CCP17 exomes were employed. Ingenuity Variant Analysis (Qiagen, Venlo, Netherlands) was used for variant annotation. Single nucleotide variations (SNVs), and small insertions and deletions were selected using following criteria: i) quality score >30; ii) at least 30X of coverage (coverage was manually checked using Integrative Genomics Viewer, IGV, https://software.broadinstitute.org/software/igv/) software if necessary; iii) MAF (Minor allele frequency) <1% in ExAC database (http://exac.broadinstitute.org/), gnomAD database (https://gnomad.broadinstitute.org/), 1000 Genomes Project database (https://www.internationalgenome.org/), and dbSNP database (https://www.ncbi.nlm.nih.gov/snp/); iv) homozygous count <3 in ExAC and gnomAD databases; v) Nonsynonymous variants in coding or splicing regions (synonymous, intronic, and other non-coding variants were also considered only provided their already described pathogenic role).

SureCall and IGV were used for coverage analysis to detect large homozygous deletions in genes known to be prone to these rearrangements.

Variant classification was made according to American College of Medical Genetics and Genomics published guidelines (Richards et al., 2015). To define the impact of missense mutations on protein function, we used an *in silico* pipeline encompassing eleven prediction tools including SIFT (https://sift.bii.a-star.edu.sg/), PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/), MutationTaster (http://www.mutationtaster.org/), FATHMM-MKL (http://fathmm.biocompute.org.uk/fathmmMKL.html), MetaSVM (https://varsome.com/), MetalR (https://varsome.com/), GERP (https://varsome.com/), LRT (https://varsome.com/), Mutation Assessor (http://mutationassessor.org/r3/), Provean (http://provean.jcvi.org/index.php) and CADD (https://cadd.gs.washington.edu/). Splicing variants and synonymous variants close to splicing sites were also tested using Human Splicing Finder 3.1 (http://www.umd.be/HSF/), NNSPLICE 0.9 (http://www.fruitfly.org/seq_tools/splice.html) and NetGene2 (http://www.cbs.dtu.dk/services/NetGene2/).

3.1.5 Sanger sequencing

Regions containing selected variants of interest were amplified by polymerase chain reaction (PCR). PCR products were purified with ExoSAP-IT™ PCR Product Cleanup Reagent (Thermo Fisher Scientific) and sequenced using BigDye™ Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific) through automatized capillary sequencing 3500xL Genetic Analyzer (Thermo Fisher Scientific). Electropherograms were analyzed using SeqScape™ Software v3.0 (Thermo Fisher Scientific).

3.1.6 Multiplex Ligation-dependent Probe Amplification (MLPA) analysis

Multiplex Ligation-dependent Probe Amplification (MRC-Holland, Amsterdam, Netherlands) was performed to determine deletions/duplications, in patients harboring a single mutation in a frequent recessive gene in order to find second mutations. We used Salsa kits P213 for *SPG7*, P441 for *SACS*, and P163 for *WFS1*. Capillary sequencing for MLPA products was performed using 3130xl Genetic Analyzer (Thermo Fisher Scientific), and Coffalyser software (MRC-Holland) was employed to analyze MLPA results.

3.2 Studies in skin fibroblasts

3.2.1 Cell cultures

Primary cultures of human skin fibroblasts were obtained from skin biopsies of affected patients and family members, when available. Primary fibroblast cell lines were grown at 37°C with 5% CO_2 in Dulbecco's modified Eagle's medium (DMEM), containing 10% fetal bovine serum (FBS), 4.5 g/L glucose and 1% antibiotics/antimycotics.

3.2.2 Quantitative real time PCR

RNA was obtained from cultured skin fibroblasts using High Pure RNA Isolation Kit (Roche), 1μg of total RNA was retro-transcribed using Transcriptor First Strand cDNA Synthesis Kit (Roche) following manufacturer's instructions. qPCR was performed using qPCRBIO SyGreen Mix Hi-ROX (PCR Biosystems, Wayne, PA.) and 7500 Fast Real-Time PCR Systems (Thermo Fisher Scientific,). The ribosomal protein 18S was used as internal control (Periasamy et al. 2010) (De Marco et al., 2016). Gene expression levels were referred to the internal control, the relative

quantification was carried out by means of the ΔCt method and the results were expressed as relative mRNA expression. Primers used are listed in supplemental material (Table S1).

3.2.3 Western blot

Cultured skin fibroblasts were collected at confluence, washed twice with PBS and then homogenized on ice using M-PER™ Mammalian Protein Extraction Reagent (Thermo Fisher Scientific) and Halt™ Protease Inhibitor Cocktail (100X) (Thermo Fisher Scientific).

Homogenates were centrifuged for 10 min at 14,000 g at 4 °C, then supernatants were collected and either stored at 4°C for short-term use or at -20 °C for long-term storage. 30-50 μg of proteins were loaded in Novex™ 8-16% Tris-Glycine Mini Gels, WedgeWell™ format (Thermo Fisher Scientific) using Novex™ Tris-Glycine SDS Sample Buffer (2X) (Thermo Fisher Scientific) and NuPAGE™ Sample Reducing Agent (10X) (Thermo Fisher Scientific). Blotting used Trans-Blot Turbo Transfer System (Bio-Rad Laboratories Inc, Berkeley, CA) and Trans-Blot® Turbo™ RTA Mini PVDF Transfer Kit (Bio-Rad). Membranes were blocked with TBS/0.1% Tween20 (TTBS) containing 5% non-fat dry milk.

Primary antibodies were incubated overnight at 4°C in TTBS with 2.5% non-fat dry milk, whereas secondary antibody peroxidase-conjugated anti-mouse and anti-rabbit were used (Jackson ImmunoResearch, laboratories Inc.) for 1 hour at room temperature in the same buffer used for the primary antibodies (2.5% non-fat dry milk in TTBS). Bands were revealed using Clarity Western ECL Substrate (Bio-Rad) and ChemiDoc™ Imaging System (Bio-Rad). ImageJ software (https://imagej.nih.gov/ij/) was used for densitometry analysis.

Primary antibodies were used as follows: rabbit polyclonal anti-AFG3L2 antibody 14631-1-AP (Proteintech, Rosemont, IL) diluted 1:500, mouse monoclonal anti-SPG7 antibody OTI1C1 (Thermo Fisher Scientific) diluted 1:1000, rabbit monoclonal anti-HARS antibody ab140640 (Abcam, Cambridge, UK) diluted 1:500, rabbit polyclonal anti-COQ4 antibody PA5-60171 (Thermo Fisher Scientific) diluted 1:1000, mouse monoclonal anti-GAPDH antibody ab8245 (Abcam) diluted 1:8000, polyclonal rabbit antibody anti-β-tubulin (Cell Signaling Technology, Danvers, MA) diluted 1:1000.

3.2.4 Micro-oxygraphy analysis

Oxygen consumption rate (OCR) was measured in fibroblasts, using an XFe24 Extracellular Flux Analyzer (Agilent Technologies) (Figure 11). Cells were plated in XF 24-well cell culture microplates at a density 6E+05 cells/well in 250 μL of normal culture media and incubated for 24h at 37°C in 5% CO_2 atmosphere. Subsequently, growth medium was replaced by 500 μL of XF medium (non-buffered DMEM medium, containing 10mM glucose, 2 mM pyruvate, 2 mM glutamine; pH 7.4), and plates were warmed at 37°C for 30 minutes. Cartridges were loaded consecutively with 1 μg/ml oligomycin (Sigma-Aldrich Corporation) to measure ATP production and proton leak after determination of basal respiration, 1.5 μM carbonyl cyanide 4-(trifluoromethoxy)phenylhydrazone (FCCP) (Sigma-Aldrich) to measure the maximal respiration, and with 0.5 μM of both mixed antimycin A (Sigma-Aldrich) and rotenone (Sigma-Aldrich) to measure non-mitochondrial oxygen consumption under the total inhibition of the respiratory chain (Invernizzi et al., 2012).
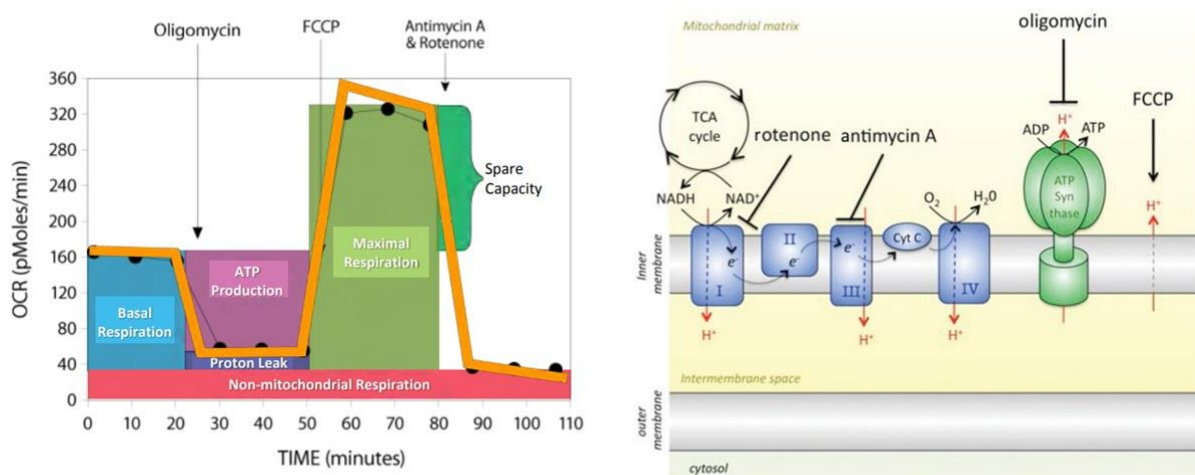
Figure 11. A schematic overview of the mitochondrial stress test using the extracellular flux analyzer. Oxygen (O_2) consumption rate (OCR) is measured prior to the addition of drugs (basal OCR) and then following the addition of the indicated drugs. Reduction in OCR after oligomycin (an ATP synthase inhibitor) indicates the amount of O_2 consumed for mitochondrial ATP production, whereas proton leak is the ratio between basal OCR and O_2 used for mitochondrial ATP generation. FCCP (a protonophore that uncouples ATP synthesis from oxygen consumption) allows H^+ back into the matrix independent of the ATP synthase; cells attempt to maintain the chemiosmotic gradient after FCCP by moving H^+ back out to the intermembrane space, which requires the use of the electron transport chain (ETC) and the consumption of O_2 as the final electron acceptor. After FCCP the maximum capacity of the mitochondria to use OXPHOS is revealed. Spare capacity is the ratio between maximal OCR and basal OCR and as such is an indicator of how close to its bioenergetic limit the cell is functioning (Nicholls et al., 2010). Rotenone (a Complex I inhibitor) and Antimycin A (a Complex III inhibitor) together render a complete shutdown of the ETC, and thus mitochondrial O_2 consumption (left figure from www.agilent.com, right figure from Van der Windt et al., 2016).

CyQUANT Cell Proliferation Assay (Thermo Fisher Scientific) was used to determine DNA amount in each well for subsequent normalization performed using Seahorse Bioscience Report Generators software (Agilent Technologies).

3.2.5 Aminoacylation assay

Measurement of the aminoacylation activity (Figure 12) was assayed using established methods (Zhang et al., 2014). Whole cell extracts were prepared from control or patient-derived fibroblasts after washing twice with cold PBS by lysing cells in 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 5 mM DTT, 0.5% Triton X-100, and protease inhibitor cocktail (Sigma-Aldrich). Protein concentration was measured using standard Bradford assay.

Aminoacylation assays were performed at 37 °C in 100 mM HEPES, pH 7.2, 30 mM KCl, 10 mM MgCl2 with 5 µM tRNAHis, 2 mM ATP, 50 µM [^{14}C] histidine, and the reaction was initiated by adding protein extract to achieve a final concentration of 0.3 µg/µL of total protein. At three different time points over a 10-min interval, 5 µL aliquots were spotted onto 3MM Whatman filter papers presoaked with 5% trichloroacetic acid (TCA). The dried paper was washed three times with 5% TCA, once with 95% ethanol, and radioactivity was quantitated by liquid scintillation counting. To calculate the specific activity of each sample, aminoacylation rates (pmol aminoacylated tRNA/min) was calculated from linear fits of the progress curve data.

Figure 12. Scheme of the two-step aminoacylation reaction. The aminoacyl-tRNA synthetase (ARS; purple), amino acid (blue), ATP, AMP, pyrophosphate (orange), and tRNA (green) are all indicated. (continue in the next page)

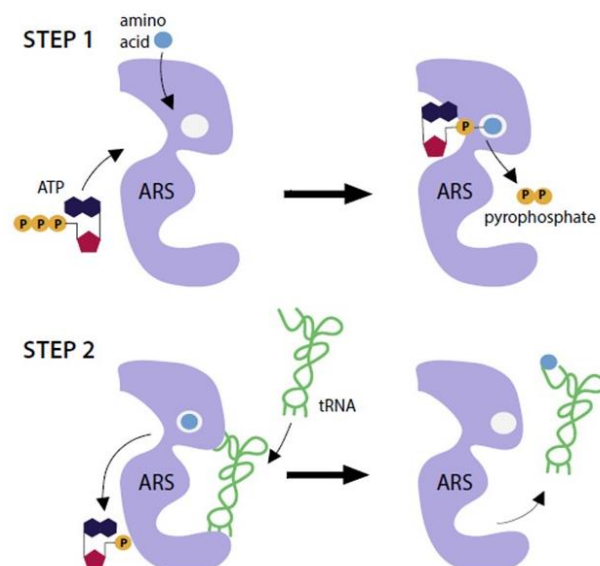The reaction occurs in two steps, as depicted. First, the ARS binds the amino acid and ATP to form the amino-adenylate intermediate, which releases pyrophosphate. In the second step, the ARS binds the cognate tRNA to facilitate transfer of the amino acid to the tRNA. The tRNA is then released for protein synthesis. The image represents a monomeric enzyme; however, the reaction proceeds similarly for oligomeric enzymes. Aminoacylation assays are used to determine the ability of mutant ARS enzymes to charge tRNA *in vitro* compared to the wild-type enzyme. Purified recombinant ARS enzyme (wild-type or mutant) is incubated with tRNA, ATP, and radiolabeled amino acid. Aliquots of the reaction mixture are collected, spotted on filter paper, and the tRNA is precipitated using trichloroacetic acid to remove any unincorporated amino acid. Radioactivity levels are then assessed to determine the amount of amino acid ligated to tRNA molecules. Steady-state kinetics are calculated by fitting the initial rate of aminoacylation as a function of tRNA concentration (or the concentration of another substrate) to the Michaelis–Menten equation (from Oprescu et al., 2017).

3.3 Studies in *Saccharomyces cerevisiae*

3.3.1 Yeast complementation assay

Yeast complementation assays were performed as described in detail by Vester and collaborators (Vester et al., 2013). Briefly, mutation-containing oligonucleotides were designed to model several mutations we identified as potentially disease-causing. The QuickChange® II XL Site-Directed Mutagenesis Kit (Stratagene, San Diego, CA.) was used (following manufacturer's instructions) to mutate every locus of interest in a pDONR221 Gateway® entry clone (Invitrogen). Resulting clones were purified and sequenced to confirm successful mutagenesis and exclude polymerase-induced mutations. Each mutated entry clone was subsequently recombined into a Gateway®-compatible LEU2-bearing pRS315 destination vector. Resulting clones were purified and digested with BsrGI (New England Biolabs, Ipswich, MA) to confirm successful recombination.

For each mutation of interest, two independently generated haploid strains (harbouring a pRS316 maintenance vector to express wild-type gene of interest and URA3) were transformed with a LEU2-bearing pRS315 vector containing no insert or containing a wild-type or mutant allele. Subsequently, yeast strains were selected on medium lacking uracil and leucine (Teknova, Hollister, California, U.S.) to select for the presence of both vectors. For each transformation, four colonies were grown to saturation in selective medium for 48 h. Next, 10 μl of undiluted and diluted (1:10 and 1:100) samples from each culture were spotted on plates containing 0.1% 5-fluoroorotic acid (5-FOA) or SD -leu -ura growth medium (Teknova, Hollister, California, U.S.) and incubated at 30°C for 48 h.

Survival was determined by visual inspection of growth. Experiments were performed using two independently generated expression constructs for each allele.

3.4 Studies in *Drosophila melanogaster*

3.4.1 *Drosophila* stocks and maintenance

Fly stocks were maintained at 25°C on standard cornmeal-agar medium (water: 36 l; yeast: 720 g; cornmeal: 3200 g; soy flour: 400 g; light malt extract: 3200 g; sugar beet syrup: 880 g; agar: 320 g, Nipagin 120 g). The crosses between the GAL4 drivers and the UAS responder lines were carried out at 25°C unless stated. UAS constructs were used in heterozygous configurations for all experiments. GAL4 drivers and UAS responders fly lines used are listed in the table. In all experiments, flies were maintained in standard condition unless stated differently. Flies were raised at 25°C with 12h:12h light/dark cycle and collected within 48h post-eclosion. CO_2 was used for anesthesia. UAS/GAL4 system has been used for targeted gene expression of fly lines triggering RNA interference of the two genes in the whole fly (*actin*-GAL4, Bloomington Stock 3954) and in a tissue specific manner (*Elav*-GAL4, Bloomington Stock 458, all neurons; *Mef2*-GAL4, Bloomington Stock 27390, muscles; *MB*-GAL4, Bloomington Stock 854 mushroom bodies; *Repo*-GAL4, Bloomington Stock, all glia). UAS-*mcherryRNAi* (Bloomington Stock 35787) was used as control line in all experiments. Mitochondrial morphology was examined using a special form of GFP targeted to this organelle (UAS-*mitoGFP* Bloomington Stock 8442, mitochondria). Two UAS lines expressing the RNAi for *SPG7* Drosophila orthologue (*dSPG7*) were available (UAS-*dSPG7RNAi*, Bloomington Stock 55324; UAS-*dSPG7RNAi*, Bloomington Stock 57843), and one for *AFG3L2* Drosophila orthologue (*dAFG3L2*) RNAi (UAS-*dAFG3L2RNAi*, Bloomington Stock 50524). UAS-*dSPG7* Bloomington Stock 32597 was also used for *dSPG7* overexpression, and *Tub*-GAL80^ts Bloomington Stock 7108 for silencing of GAL4-UAS system.

3.4.2 Quantitative real time PCR

Total RNA was extracted from 15 whole flies using peqGold TriFast reagent (PEQLAB Biotechnologie GMBH, Erlangen, Germany). 500 ng mRNA were retro-transcribed into cDNA using QuantiTect Rev. Transcription Kit (Qiagen GmbH, Hilden, Germany) and then employed for qPCR with ORA qPCR Green ROX L Mix (HighQu, Kralchtal, Germany) on a CFX connect™ Real-time PCR Detection System (Bio-Rad).

The ribosomal protein 49 (rp49) was used as internal control. Results were analyzed with the Bio-Rad CFX manager3.1 software (Bio-Rad) Gene expression levels were referred to the internal control, the relative quantification was carried out by means of the Ct method and the results were

expressed as relative mRNA expression. Primers used are listed in supplemental material. (Table S1).

3.4.3 Lifespan assay

Flies longevity was measured using at least 100 individuals per genotype. Experiments were conducted using a starting density of 25 flies per vial, and using standard cornmeal agar medium. Every 2-3 days, dead flies were counted and survivals were carried in a new vial with fresh food. No anesthesia was used during experiment. Median lifespan (MedLs) and maximum lifespan (MaxLs) was indicated for each experiment.

3.4.4 Negative geotaxis assay

Negative geotaxis was tested to assay flies locomotion. Negative geotaxis is the measure of how fast flies are able to climb vertically after being blown at the bottom of a narrow vial that do not let flies to fly, hence testing their innate escape response (Linderman et al., 2012). The distance that a fly can climb in 12 seconds was measured using clean vial. No anesthesia was used in transferring flies in testing vials. 12-15 flies per genotype were tested and each fly was recorded three times and the mean value of each fly was used for the subsequent analysis.

3.4.5 Brain histology

Brain histology studies were performed as described elsewhere (Edenharter et al., 2017). To determine the brain integrity, paraffin section were performed fixing anesthetized flies with carnoy (ethanol:chloroform:acetic acid at a 6:3:1 proportion), dehydrated with pure ethanol and then embedded in paraffin. About 7 μm paraffin brain sections were analyzed using a fluorescence microscope. Brain vacuolization was checked in the total brain area, including the lamina, the outer chiasm and the medulla.

For immunofluorescence assays, whole brains were extracted and analyzed. Flies were fixed using 4% paraformaldehyde (PFA) for 2h. Then, brains were dissected, washed with PBST and blocked using 10% Normal Goat Serum. Primary antibodies were incubated 24h at 4°C, washed with PBST and then incubated 24h at 4°C with secondary antibodies. Samples were then mounted using Vectashield mounting medium (Vector Laboratories, Burlingame, CA.) and scanned using multi-

tracking setups of either LSM 510 Confocal Microscope (Oberkochen, Germany) or Confocal Laser Scanning Platform Leica TSC SP8 (Wetzlar, Germany).

 Following antibodies were used: mouse anti-FasII (1:10, labels mushroom bodies) and mouse anti-Bruchpilot (1:10, labels all neurons) as primary antibodies, both from Hybridoma Bankand. Goat anti-mouse AlexaFluor 555 (Thermo Fisher Scientific) was used as secondary antibody in both cases.

Schematic figures of fruitfly brain structures are shown in figure 13A, 13B, 13C and 13D.
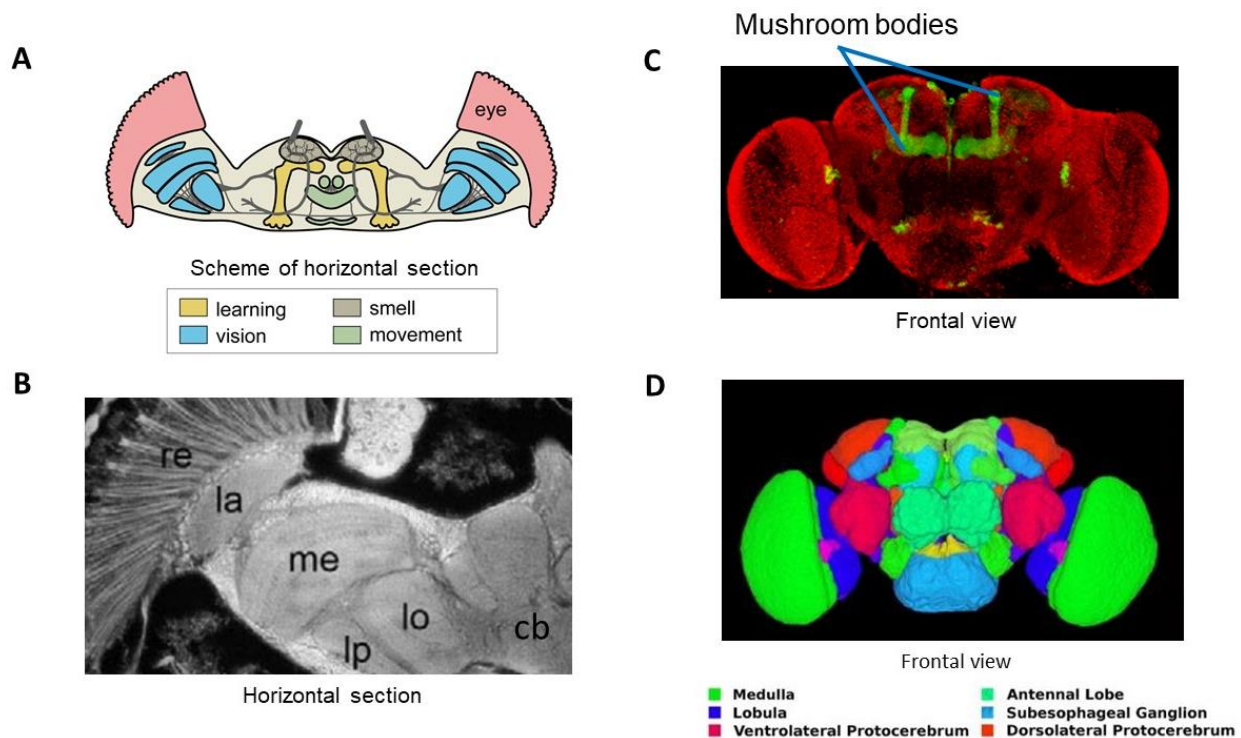
Figure 13. Anatomical and functional organization of Drosophila brain. (A) Schematic horizontal section showing functional organization of fruit fly brain. In Drosophila, a consistent part of the brain has a role in vision. (B) Horizontal section indicating anatomical organization of Drosophila brain mainly involved in vision. Legend: re=retina, la=lamina, me=medulla, lo=lobula, lp=lobula plate, cb=central brain. (C) Frontal view of the entire fruit fly brain; green signal indicates labelled mushroom bodies, red signal denote the rest of neurons. (D) Anatomic organization of the whole brain. Medulla and lobula are involved in vision (Zhu, 2013), antennal lobe in olfactory system (Kim et al., 2015), subesophageal ganglion in processing of gustatory information and the control of feeding behaviour (Hartenstein et al., 2018), protocerebrum in higher sensory processes (Lai et al., 2012) (part A is from droso4schools project of Manchester Fly Facility, https://droso4schools.wordpress.com/organs/, image of Andreas Prokop; part B is modified from Dutta et al., 2016; part C is from Fitzsimons and Scott, 2011; part D credit: Copyright 2006-2015 Hanchuan Peng and HHMI - Janelia Research Campus, image distributed with permission under Vaa3D License, https://medicalxpress.com/news/2015-12-collaborative-brain.html).

3.4.6 Muscle histology

To evaluate the integrity of both muscle fibers and mitochondria ultrastructure, we studied flight muscles. Thoraxes were isolated from anesthetized flies and fixed as described for brain dissection. Thoraxes were then sagitally cut and incubated 24h at 4 °C with phalloidin to visualize F-actin (Sigma-Aldrich) from muscle fibers, whereas fluorescence from a mitochondrially-targeted version of GFP was used to visualize mitochondria. Samples were mounted and observed as previously described.

3.4.7 Measurement of aconitase activity

Aconitase activity was measured using Bioxytech Aconitase-340^{TM} Spectrophotometric Assay kit (OXIS International Inc., Portland, OR). 25 adult flies (n= 25) or thoraxes (n= 15) obtained from anesthetized flies were squished in 500 ml of ice-cold homogenization buffer (0.2 mM sodium citrate and 50 mM Tris-HCl, pH 7.4). Homogenates were then centrifuged twice at 800 g for 10 minutes at 4°C, supernatants were recovered every time and disrupted by sonication for 1 minute, with 1 pulse every 2 seconds.

The aconitase activity of each sample was then related to its protein amount, measured using the Bradford assay (Coomassie PlusTM Protein Assay Reagent, Thermo Fisher Scientific) in a Tecan Infiniter F500 plate reader (Tecan Group Ltd, Zürich, Switzerland). Each experiment consisted of 3-5 independent biological replicates.

3.4.8 Measurement of ATP production

ATP rate production was measured using ATP Bioluminescence Assay Kit CLS II (Roche). Full flies (n= 15) or thoraxes (n= 5) obtained from anesthetized flies were homogenized in 200 µl pre-heated ATP assay buffer (100 mM Tris, 4 mM EDTA, pH 7.75), boiled for 2 min at 95°C and centrifuged for 1 min at 1000 g. The supernatant was diluted 1:3 in ATP assay buffer and 50 µl of the dilution were used to measure ATP levels. Luminescence was detected with a Tecan Infiniter F500 plate reader (Tecan Group). Each experiment consisted of 3–5 independent biological replicates.

3.4.9 Hyperoxia conditions

Oxidative stress induced by hyperoxia is a valid tool to identify novel roles involved in neurodegenerative disorders (Gruenewald et al., 2009), so it was coupled to some of the assays previously described. Treatment in hyperoxia condition begun 1 day post-eclosion and was performed by exposing flies in a glass container with 99.5% oxygen constant flux and under a low positive pressure at 25°C. Flies were carried in new vials with fresh food every day.

3.4.10 Generation of a *SPG7* knock-out model

A *SPG7* KO model was generated using CRISPR/Cas9 technology. Two different guide RNAs (gRNA) were designed and cloned in two different vectors, and two different regions (that we called "homology arms" and cloned in a third vector containing a red fluorescent eye marker) were chosen for insertion of a cassette in *SPG7* Drosophila orthologue gene (*dSPG7*) through homologous recombination to cause full gene loss of function. The three plasmids were then injected in Drosophila embryos constitutively expressing Cas9 as already described (Gratz et al., 2015), and the fluorescent marker was used to select the transformant offspring.

gRNAs were created through oligo annealing using 10x T4 ligation buffer and T4 Polynucleotide Kinase (both New England Biolabs) to phosphorylate the oligos (thermocycler program: 37°C for 30 min, 95°C for 5 min, then ramp to 25°C at a rate of -0.1°C/sec). pU6-BbsI-chiRNA (Addgene, Plasmid #45946) was then digested with BbsI restriction enzyme (New England Biolabs) 4 hours at 37°C. BbsI digested pU6 and annealed oligo inserts were then ligated using T4 DNA ligase (New England Biolabs.) 18°C overnight (two different reactions for the two different gRNAs). Plasmid containing gRNA were transformed using DH5-Alpha *E.coli* cells in LB/ampicillin selective medium. Primers used are listed in supplemental material. (Table S1).

Two *SPG7* regions in the Drosophila genome were amplified to create the homology fragments necessary for homologous recombination with Drosophila genome, and to create also specific enzyme restriction sites. AgeI and SacII restriction enzymes were used to digest *dSPG7* homology arm 1 (*dSPG7*hom1), whereas KpnII and SpeI for *dSPG7* homology arm 2 (*dSPG7*hom2) (all from New England Biolabs), all 4 hours reaction at 37°C. pT-GEM plasmid (Addgene, Plasmid #62893) was then digested with same restriction enzymes used for *dSPG7*hom1 (AgeI and SacII). Then, pT-GEM and *dSPG7*hom1 were ligated using T4 DNA ligase (New England Biolabs.) 18°C overnight. Plasmid containing gRNA were transformed using DH5-Alpha *E.coli* cells in LB/ampicillin selective medium. Once obtained, the pT-GEM/*dSPG7*hom1 was digested with the same restriction

enzyme used for *dSPG7*hom2 (namely, *Kpn*II and *Spe*I). pT-GEM/*dSPG7*hom1 and *dSPG7*hom2 were then ligated and transformed as described for *dSPG7*hom1. Primers used are listed in supplemental material. (Table S1).

Both plasmid containing the different gRNAs, and pT-GEM/*dSPG7*hom1/*dSPG7*hom2 plasmidic construct were then purified using Qiagen Plasmid Mini Kit (Qiagen) for Sanger sequencing analysis to check the correct cloning processes, whereas Qiagen Plasmid Midi Kit (Qiagen) was employed to collect a large amount of purified plasmid for subsequent injection in Drosophila embryos.

Embryos from transgenic flies expressing Cas9 were injected with a solution containing both vectors carrying the gRNAs and pT-GEM/*dSPG7*hom1/*dSPG7*hom2 vector, survived flies were crossed to white mutant flies and the offspring of those crosses was screened to identify flies containing the red fluorescent marker 3xP3 in the eye, coded by pT-GEM vector. Transformant lines were genotyped by Sanger sequencing to check the correct and precise insertion.

3.4.11 Setting up a new *Spg7^Ala510Val* knock-in model

For preparation of fly knock-in model of paraplegin, two novel guide RNAs (gRNA) and a *de novo* vector, containing the *dSPG7* sequence harboring the p.Ala510Val mutation to insert, were created. Primers used are listed in supplemental material (Table S1).

New vectors containing new gRNAs were created exactly in the same way as previously described, but using new primers.

The new vector designed for introduction of p.Ala510Val mutation was assembled through Gibson Assembly (New England Biolabs). Fragment conferring origin of replication site and resistance to ampicillin was amplified from standard pT-GEM vector, whereas fragments necessary for the new exchange with the sequence harboring Ala510Val variant were amplified using Drosophila KO genomic DNA. Total amount of each fragments and thermocycler conditions were used as described by manufacturer's instruction. Transformation in DH5-Alpha *E.coli* cells, purification of plasmid for sequencing and for injection were performed as previously described.

Line for injection was prepared crossing *SPG7* KO flies and transgenic flies expressing Cas9.

3.5 Studies in *Danio rerio*

3.5.1 Zebrafish maintenance

Adult male and female wild-type (WT) zebrafish (AB strains) and transgenic line expressing GFP under a strong neuronal promoter, Tg(Neurod1-GcAMP6f) were maintained according to standard procedures (Westerfield, 2000) on a 14 hour light:10 hour dark cycle. Fish system water conditions were kept in the following ranges by automated feedback controls: 28.5°C, pH 7.5-8.0, conductivity 690-710 mS/cm. Zebrafish embryos and larvae were raised in an incubator kept at 28.5°C, on the same light/dark cycle as the fish facility. Water used for embryos and larvae was prepared by adding 0.03% Instant Ocean and 0.000002% methylene blue to reverse osmosis-treated water.

All animal experiments were carried out in accordance with the EU Directive 2010/63/EU for animal experiments and all procedures complied with the guidelines of our institutional animal care committee. Zebrafish embryos and larvae procedures complied with the guidelines of our institutional animal care committee, and experiments were performed in accordance with, and under the supervision of, the Institutional Animal Care and Use Committee (IACUC) of the University of Pisa. Every effort was made to minimize both animal suffering and the number of animals needed to collect reliable scientific data.

3.5.2 mRNA overexpression

For the overexpression experiments, total RNA was extracted using the Quick RNA miniprep (Zymo Research) from embryos at 48 hours post fertilization (hpf) and reversely transcribed using the Transcriptor First Strand cDNA Synthesis Kit (Roche, Basel, Switzerland). Primers for mRNA sequences were designed using the zebrafish sequence of coq4 (ENSDART00000164067.3), and the human transcript (ENST00000300452.8). PCR products were cloned into pCS2+ vector and verified for accuracy using capillary sequencing. The overexpression was performed through co-injection of 150 pg of control, either human WT or mutant complementary RNAs (cRNAs). Each experiment was repeated at least three times if not otherwise stated.

3.5.3 Behavioural and locomotor studies

Survival was calculated comparing death rates between microinjected and non-injected embryos at 24 hpf.

Touch-evoked escape response was measured at 48 hpf on a semiquantitative scale from severe, no movement; mild, flicker of movement but no swimming; to wild-type, normal swimming.

Coiling behavior was measured in embryos at 30 hpf (n=20 for each experiment) using the Danioscope software (Noldus Information Technology, Wageningen, The Netherlands).

Each experiment was repeated three times, while the locomotor tracking of 5dpf larvae was measured on morphants and control larvae (n=24 for each experiment) using the Daniovision system and the Ethovision XT software (Noldus).

3.5.4 Quantitative real time PCR

Quantitative real rime PCR was performed as described in section 3.3.2. β-actin was used as internal control. Primers used are listed in supplemental material (Table S1).

3.5.5 Whole-mount *in situ* hybridization

Whole mount in situ hybridization (WISH) was performed as described in detail by Thisse and Thisse (Thisse and Thisse, 2008) at the 48, 72, and 96 hours post fertilization (hpf) time points. Briefly, the sequence of the gene of interest is amplified by PCR, and used as a template for the synthesis of an antisense RNA probe, which is labeled with digoxigenin-linked nucleotides using the Digoxigenin (DIG) RNA Labeling Kit (Roche). Embryos are fixed and permeabilized before being soaked in the digoxigenin-labeled probe, conditions that favor specific hybridization to complementary mRNA sequences in the tissues expressing the corresponding gene. After washing away excess probe, hybrids are detected by immunohistochemistry using an alkaline phosphatase-conjugated antibody against digoxigenin and a chromogenic substrate.

3.5.6 Whole-mount immunohistochemistry

Whole-mount immunohistochemistry, embryos were fixed in 4% PFA overnight at 4°C and stored in methanol. Anti-pvalb7 (1:1000) antibody was used to stain Purkinje cells, the antibody was a gift of Hibi Lab and performed as previously described (Bae et al, 2009). Anti-tubulin acetylated antibody (1:1000) (Sigma-Aldrich, clone 6-11B-1) was used to label motorneurons. Images of the brain defects were acquired using the stereomicroscope Leica M205FA (Leica, Wetzlar, Germany). Scheme and example of zebrafish larval brain is shown in Figure 14A and 14B.
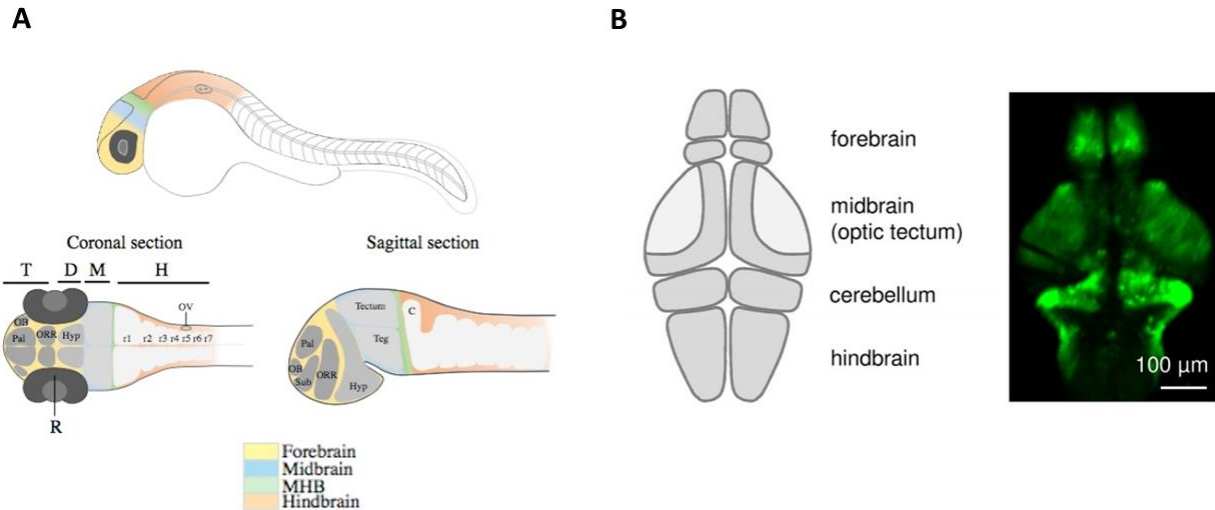
Figure 14. Zebrafish larval brain. (A) Schematic representation of the embryonic brain (30 hours post fertilization, hpf), showing the forebrain (in yellow), midbrain (in blue), MHB (in green), and hindbrain (in orange). Coronal and sagittal section schemes show brain structures primordia. Forebrain is subdivided in the telencephalon (in darker grey) and the diencephalon (containing the hypothalamus, lighter grey). C: cerebellum; D: diencephalon; M: midbrain; MHB: midbrain-hindbrain boundary; H: hindbrain; Hyp: hypothalamus; OB: olfactory bulb; ORR: optic recess region; Pal: pallium; R: retina; r1–r7: rhombomeres 1 to 7; Sub: sub-pallium; T: telencephalon; and Teg: tegmentum. (B) Schematic illustration (left) and image (right) depicting sub-regions of the larval zebrafish brain (part A is modified from Vaz et al., 2019; part B is from Liu and Baraban, 2019).

3.5.7 d-Cas9 knock-down model

For the generation of zebrafish *coq4* knockdown, we used the CRISPR interference (CRISPRi) method. Dead-Cas9 (dCas9) was a kind gift of prof. Andrea Rossi (Bonn, Germany) and it has already been published (Rossi et al., 2015). CRISPR gRNAs were designed using CHOPCHOP (https://chopchop.cbu.uib.no) (Primers used are listed in supplemental material, Table S1). To make dCas9 RNA, the template DNA was linearized by XbaI digestion and purified using DNA clean and contractor KIT (Zymo Research, Irvine, CA), and RNA was synthesized using a mMESSAGE mMACHINE T3 kit (Life Technologies, Carlsbad, CA) and purified using an RNA Clean and Concentrator kit (Zymo Research). To make sgRNA, the template DNA obtained through the annealing of oligonucleotides was purified using a DNA clean and contractor KIT (Zymo Research), and after was in vitro transcribed using a T7 RNA polymerase MEGA short script T7 kit (Life Technologies). After in vitro transcription, the sgRNA (about 100 nucleotides long) was purified using RNA clean and concentrator (Zymo Research,). dCAS9 mRNA (150 pg) and gRNA (50 pg) were co-injected in the cell at the one-cell stage and at least ten pooled embryos were used to evaluate the expression level of the targeted genes by qPCR.

3.5.8 Generation of a *coq4* knock-out model

To generate CRISPR/cas9 knock-out model, we used the pCS2-Cas9 plasmid to generate Cas9 mRNA (http://www.addgene.org/Alex_Schier/), which was transformed in liquid LB/ampicillin under standard conditions. DNA was prepared using a standard miniprep protocol, mixed and incubated for 1 hour at 37˚C, then purified using standard DNA column purification protocol. The linearized pCS2-Cas9 plasmid was transcribed capped Cas9 mRNA using mMessage mMachine SP6 polymerase Transcription kit (Ambion, Life Technologies), incubated at 37˚C for 4 hours, and then 1 µl of TURBO DNase was added and incubate for 15 minutes at 37˚C. Sequence-specific sgRNA template plasmids were generated for each target using CHOPCHOP software (Primers used are listed in supplemental material, Table S1). sgRNAs were transcribed using MEGAscript T7/SP6 (Ambion), and purified MegaClear Kit (Ambion).

Fertilized 1–2 cell stage zebrafish eggs were injected with an injection mix containing approximately 300 ng/µl Cas9 mRNA and 15 ng/µl sgRNA. After injected eggs were incubated for one day, and some were harvested to check for mutagenesis at the target site. DNA was extracted from pools of 10 injected embryos and uninjected controls, gDNA including the target site was amplified, and the target site was checked via Sanger sequencing. Multiple sequencing peaks were confirmed to be present at the sgRNA target site before proceeding.

3.6 Statistical analyses

All data are shown as mean ± SEM. At least three replicates were used in all experiments. Statistical analyses were carried out using Prism version 7.04 (GraphPad Software, La Jolla, CA, www.graphpad.com)

Unpaired t-test or one-way ANOVA with Dunnett's multiple comparison test were used for statistical analysis of quantitative real time PCR and Western blots in skin fibroblasts, and for negative geotaxis assays and measurements of aconitase activity and ATP levels in fruit fly. Multiple t-test was used for micro-oxygraphy analysis in skin fibroblasts. Survival data in fruit fly were analyzed using the log-rank (Mantel–Cox) and Gehan–Breslow–Wilcoxon tests. For studies in zebrafish, homogeneity of variance was assessed using the Levene test; Chi-square and Fisher's exact test were used for survival count and touch-evoked response assays, whereas Mann-Whitney test with Bonferroni's correction was performed for coiling frequency test and locomotion assays. No statistical analyses were employed for comparisons of mRNA levels downregulations in fruit fly

and zebrafish, since empirical evidences in these cases were employed to properly evaluate the result.

Prism p-value format was used to show significant statistical differences (*=P<0.05, **=P<0.01, ***=P<0.001, ****=P<0.0001) unless differently specified.

4. Results

4.1 Massive parallel resequencing results

We analyzed a total of 259 index cases (138 males, 53.3%; 121 females, 46.7%). Table 1 summarizes the genetic results. Sequencing coverage was significatively high, indicating very high reliability of our results (minimum number of reads/% of analyzable target regions covered: 1/99.95%, 10/99.79%, 20/99.61%, 50/99.02%, 100/96.96%, 200/86.97%, 500/80.34%, 1000/51.99%).

Among our cohort, 44/259 patients (17%) were familial cases (20 autosomal dominant and 24 autosomal recessive), whilst the great majority (215/259, 83%) were sporadic. An overall of additional 210 affected or unaffected relatives were available for further genetic tests, and at least 1 relative could be tested when examining 107 patients (41.3%).

Our cohort was highly heterogeneous, including phenotypes presenting pure cerebellar ataxias, or spastic ataxias, or congenital ataxias, or sensory ataxias, or even ataxia with seizures/myoclonus/peripheral neuropathy. We also studied more complex syndromes with ataxias as one of the prominent features. Also, both age of latest examination and age of onset were highly variable, ranging from first years of life to late adulthood. The high level of heterogeneity reflects the routine referral of highly polymorphic presentations to a third-level, NGS-based clinical neurogenetic laboratory in our national health system.

Pathogenic mutations were identified in 65 patients (25%) (Figure 15A), including 37 males (57%) and 28 females (43%). 66 patients (~25%) had an uncertain molecular diagnosis due to presence of variants of unknown significance (VOUS) and/or did not have enough phenotype data to help variants prioritization, besides the frequent impossibility to perform segregation analyses (Figure 15A). 128 patients (50%) resulted negative to TRP analysis (Figure 15A).

SCAs genes were causative in 11/65 (17%) of patients, whilst recessive genes were found to be mutated in 36/65 (55%). Interestingly, 18/65 (28%) of pathogenic mutations lay in genes not respecting zygosity rules, causing both SCAs and SCARs (Figure 15B).

Among positive cases, our variant filtering criteria defined 92 mutations as disease causing (all in silico predictions, frequencies and references are listed in Table S2 in supplemental material). Missense mutation (57/92, 62%) were the most common, whereas frameshift, nonsense, splicing, large deletion and in frame indel were less common being found in 11 (12%), 11 (12%), 8 (9%), 3 (3%) and 2 (2%) patients, respectively (Figure 15C). Large deletions could be detected due to

complete absence of coverage in the same exon in three cases, suggesting a possible homozygous exon deletion subsequently confirmed by MLPA. In cases with a positive diagnosis, we could study family relatives in 41/65 patients (64%).

Table 1. Summary of NGS results.

Index case	Gender, age	Familial history	Available samples for segregation	Gene (Ref_Seq)	cDNA variant, protein variant	Zigosity	Phenotype
Pt1	F, 36	AR	4	*AARS2* (NM_020745)	c.446G>A, p.Cys149Tyr; c.385A>C, p.Thr129Pro	Compound heterozygous	Ataxia with hypogonadism
Pt2	M, 11	SP	3	*AFG3L2* (NM_006796)	c.2167G>A, p.Val723Met; c.634dupG, p.Val212Glyfs*4	Compound heterozygous	Episodic ataxia
Pt3	F, 26	AD	NA	*AFG3L2* (NM_006796)	c.2105G>A, p.Arg702Gln	Heterozygous	Spinocerebellar ataxia 28 (SCA28)
Pt4	M, 58	AR	NA	*ANO10* (NM_018075)	c.1009T>G, p.Phe337Val; c.289delA, p.Met97*	Compound heterozygous	Spastic ataxia
Pt5	F, 43	SP	NA	*ANO10* (NM_018075)	c.206T>A, p.Leu69*	Homozygous	Spastic ataxia
Pt6	M, 2	SP	2	*ATM* (NM_000051)	c.2152dupT, p.Cys718Leufs*20; c.2929T>C, p.Cys977Arg	Compound heterozygous	Ataxia-telangiectasia
Pt7	M, 77	SP	NA	*ATP13A2* (NM_022089)	c.1205C>T, p.Thr402Met	Homozygous	Myoclonic ataxia
Pt8	F, 12	SP	2	*CACNA1A* (NM_023035)	c.4897G>A, p.Asp1633Asn	Heterozygous	Episodic ataxia type 2 (EA2)
Pt9	F, 59	AD	2	*CACNA1G* (NM_018896)	c.5144G>A, p.Arg1715His	Heterozygous	Spinocerebellar ataxia 42 (SCA42)
Pt10	M, 4	SP	2	*COQ4* (NM_016035)	c.577C>T, p.Pro193Ser; c.718C>T, p.Arg240Cys	Compound heterozygous	Congenital ataxia
Pt11	F, 20	SP	3	*COQ4* (NM_016035)	c.284G>A, p.Gly95Asp; c.305G>A, p.Arg102His	Compound heterozygous	Spastic ataxia with epilepsy
Pt12	F, 10	SP	2	*COQ8A* (NM_020247)	c.1844G>A, p.Gly615Asp; c.589-3C>G	Compound heterozygous	Recessive spinocerebellar ataxia 9 (SCAR9)
Pt13	M, 55	SP	NA	*COQ8A* (NM_020247)	c.1042C>T, p.Arg348*	Homozygous	Recessive spinocerebellar ataxia 9 (SCAR9)
Pt14	M, 17	AR	3	*EXOSC3* (NM_016042)	c.395A>C, p.Asp132Ala	Homozygous	Pontocerebellar hypoplasia type 1B (PCH1B)

44

| Index case	Gender, age	Familial history	Available samples for segregation	Gene (Ref_Seq)	cDNA variant, protein variant	Zigosity	Phenotype
Pt15	F, 48	SP	NA	*GJC2* (NM_020435)	c.254T>C, p.Val85Ala; c.219_220delCC, p.Leu74fVals*33	Compound heterozygous	Recessive spastic paraplegia 44 (SPG44)
Pt16	M, 11	SP	3	*HARS* (NM_001289094)	c.616G>T, p.Asp206Tyr; c.730delG, p.Val244Cysfs*6	Compound heterozygous	Spastic ataxia
Pt17	F, 39	AR	2	*HARS* (NM_001289094)	c.1393A>C, p.Ile465Leu; c.910_912dupTTG, p.Leu305dup	Compound heterozygous	Congenital ataxia
Pt18	M, 32	SP	3	*HSD17B4* (NM_001199291)	c.2191C>T, p.Gln731*; c.727G>T, p.Val243Leu	Compound heterozygous	Spastic ataxia
Pt19	M, 6	SP	2	*ITPR1* (NM_001168272)	c.722G>A, p.Arg241Lys	Heterozygous	Congenital ataxia
Pt20	M, 2	SP	2	*ITPR1* (NM_001168272)	c.805C>T, p.Arg269Trp	Heterozygous	Congenital ataxia
Pt21	M, 31	SP	1	*ITPR1* (NM_001168272)	c.7748T>C, p.Ile2583Thr	Heterozygous	Congenital ataxia
Pt22	M, 24	SP	2	*KCNA2* (NM_004974)	c.890G>A, p.Arg297Gln	Heterozygous	Congenital ataxia
Pt23	M, 18	SP	1	*KCNA2* (NM_004974)	c.881G>A, p.Arg294His	Heterozygous	Congenital ataxia
Pt24	F, 63	AD	NA	*KCNC3* (NM_004977)	c.1268G>A, p.Arg423His	Heterozygous	Congenital ataxia
Pt25	M, 80	SP	NA	*KCND3* (NM_004980)	c.680_682delTCT, p.Phe227del	Heterozygous	Congenital ataxia
Pt26	M, 29	SP	NA	*KIF1C* (NM_006612)	c.765delC, p.Asp256Thrfs*10	Homozygous	Recessive spastic ataxia type 2 (SPAX2)
Pt27	M, 7	SP	2	*LAMA1* (NM_005559)	c.1404_1405delAG, p.Gly469Alafs*5; c.184C>T, p.Arg62*	Compound heterozygous	Poretti-Boltshauser syndrome (PTBHS)
Pt28	F, 5	SP	2	*PLA2G6* (NM_001349864)	c.1703T>C, p.Phe568Ser; c.1111G>A, p.Val371Met	Compound heterozygous	Spastic ataxia

Index case	Gender, age	Familial history	Available samples for segregation	Gene (Ref_Seq)	cDNA variant, protein variant	Zigosity	Phenotype
Pt29	M, 33	AR	1	*PMM2* (NM_000303)	c.422G>A, p.Arg141His; c.323C>T, p.Ala108Val	Compound heterozygous	Congenital disorder of glycosylation type 1a (CGD1A)
Pt30	F, 42	AR	3	*PNPLA6* (NM_001166111)	c.1880C>T, p.Ala627Val	Homozygous	Boucher-Neuhauser syndrome (BNHS)
Pt31	M, 57	SP	1	*PNPLA6* (NM_001166111)	c.2264A>C, p.Gln755Pro; c.3388C>T, p.His1130Tyr	Compound heterozygous	Boucher-Neuhauser syndrome (BNHS)
Pt32	M, 56	AR	3	*PNPLA6* (NM_001166111)	c.3023A>G, p.Asp1008Gly; c.4075C>T, p.Arg1359Trp	Compound heterozygous	Boucher-Neuhauser syndrome (BNHS)
Pt33	M, 33	SP	2	*PNPLA6* (NM_001166111)	c.3365C>T, p.Pro1122Leu; c.4081C>T, p.Arg1361*	Compound heterozygous	Boucher-Neuhauser syndrome (BNHS)
Pt34	F, 64	AR	1	*PNPLA6* (NM_001166111)	c.3385G>A, p.Gly1129Arg	Homozygous	Boucher-Neuhauser syndrome (BNHS)
Pt35	F, 65	AR	2	*POLR3A* (NM_007055)	c.4073G>A, p.Gly1358Glu; c.1909+22G>A	Compound heterozygous	Spastic ataxia
Pt36	M, 54	AR	1	*POLR3A* (NM_007055)	c.3839dupT, p.Met1280Ilefs*20; c.1909+22G>A	Compound heterozygous	Spastic ataxia
Pt37	M, 50	AD	2	*PRKCG* (NM_002739)	c.358C>T, p.Leu120Phe	Heterozygous	Spinocerebellar ataxia 14 (SCA14)
Pt38	F, 9	SP	2	*RARS2* (NM_001318785)	c.1037C>T, p.Thr346Ile; c.517G>A, p.Asp173Asn	Compound heterozygous	Pontocerebellar hypoplasia type 6 (PCH6)
Pt39	F, 45	SP	2	*RNF216* (NM_207111)	c.1849A>G, p.Met617Val; c.2061+3A>G	Compound heterozygous	Ataxia and chorea
Pt40	F, 81	AR	NA	*SETX* (NM_001351527)	c.3242T>C, p.Phe1081Ser	Homozygous	Recessive spinocerebellar ataxia 1 (SCAR1)
Pt41	M, 52	AR	NA	*SETX* (NM_001351527)	c.7292dupA, p.Asn2431Lysfs*19	Homozygous	Recessive spinocerebellar ataxia 1 (SCAR1)
Pt42	F, 5	SP	2	*SLC2A1* (NM_006516)	c.136C>T, p.Gln46*	Heterozygous	GLUT1 deficiency syndrome
Pt43	F, 71	SP	NA	*SPG7* (NM_003119)	c.1231G>A, p.Asp411Asn; c.679C>T, p.Arg227*	Compound heterozygous	Recessive spastic paraplegia 7 (SPG7)

Index case	Gender, age	Familial history	Available samples for segregation	Gene (Ref_Seq)	cDNA variant, protein variant	Zigosity	Phenotype
Pt44	F, 57	SP	NA	*SPG7* (NM_003119)	c.1529C>T, p.Ala510Val; c.1940C>A, p.Ala647Glu	Compound heterozygous	Recessive spastic paraplegia 7 (SPG7)
Pt45	M, 72	AR	NA	*SPG7* (NM_003119)	c.1529C>T, p.Ala510Val	Homozygous	Recessive spastic paraplegia 7 (SPG7)
Pt46	M, 65	AR	2	*SPG7* (NM_003119)	c.1529C>T, p.Ala510Val	Homozygous	Recessive spastic paraplegia 7 (SPG7)
Pt47	M, 68	AR	1	*SPG7* (NM_003119)	Del exon 2	Homozygous	Recessive spastic paraplegia 7 (SPG7)
Pt48	F, 74	AR	NA	*SPG7* (NM_003119)	Del exon 2	Homozygous	Recessive spastic paraplegia 7 (SPG7)
Pt49	M, 69	AR	NA	*SPG7* (NM_003119)	c.73_80delCCAGGCCC, p.Pro25Glyfs*46; c.1940C>A, p.Ala647Glu	Compound heterozygous	Recessive spastic paraplegia 7 (SPG7)
Pt50	M, 61	AR	NA	*SPG7* (NM_003119)	Del exon 2	Homozygous	Recessive spastic paraplegia 7 (SPG7)
Pt51	F, 63	AR	NA	*SPG7* (NM_003119)	c.1529C>T, p.Ala510Val; c.1972G>A p.Ala658Thr	Compound heterozygous	Recessive spastic paraplegia 7 (SPG7)
Pt52	M, 26	SP	2	*SPTBN2* (NM_006946)	c.1438C>T p.Arg480Trp	Heterozygous	Spinocerebellar ataxia 5 (SCA5)
Pt53	M, 54	SP	NA	*SPTBN2* (NM_006946)	c.1843C>T, p.Arg615Trp; c.157+1G>A	Compound heterozygous	Recessive spinocerebellar ataxia 14 (SCAR14)
Pt54	M, 63	AD	5	*STUB1* (NM_005861)	c.97G>A, p.Gly33Ser	Heterozygous	Spinocerebellar ataxia 48 (SCA48)
Pt55	M, 52	AD	1	*STUB1* (NM_005861)	c.689_692delACCT, p.Tyr230Cysfs*9	Heterozygous	Spinocerebellar ataxia 48 (SCA48)
Pt56	M, 61	AD	4	*STUB1* (NM_005861)	c.682C>T, p.Pro228Ser	Heterozygous	Spinocerebellar ataxia 48 (SCA48)
Pt57	F, 53	AD	2	*STUB1* (NM_005861)	c.199G>A, p.Ala67Thr	Heterozygous	Spinocerebellar ataxia 48 (SCA48)

Index case	Gender, age	Familial history	Available samples for segregation	Gene (Ref_Seq)	cDNA variant, protein variant	Zigosity	Phenotype
Pt58	F, 48	SP	NA	*STUB1* (NM_005861)	c.673C>T, p.Arg225*	Heterozygous	Spinocerebellar ataxia 48 (SCA48)
Pt59	F, 55	SP	NA	*STUB1* (NM_005861)	c.721C>T, p.Arg241Trp	Heterozygous	Spinocerebellar ataxia 48 (SCA48)
Pt60	M, 52	SP	NA	*STUB1* (NM_005861)	c.170C>T, p.Pro57Leu	Heterozygous	Spinocerebellar ataxia 48 (SCA48)
Pt61	M, 32	SP	2	*SYNE1* (NM_182961)	c.15049C>T, p.Gln5017*	Homozygous	Recessive spinocerebellar ataxia 8 (SCAR8)
Pt62	M, 61	SP	NA	*SYNE1* (NM_182961)	c.7085dupA, p.Asn2369Lysfs*4; c.6724-1G>A	Compound heterozygous	Recessive spinocerebellar ataxia 8 (SCAR8)
Pt63	M, 39	AR	1	*SYNE1* (NM_182961)	c.4609C>T, p.Arg1537*	Homozygous	Recessive spinocerebellar ataxia 8 (SCAR8)
Pt64	M, 83	SP	NA	*TMEM216* (NM_001173991)	c.432-1G>C	Homozygous	Joubert syndrome 2 (JBTS2)
Pt65	F, 74	SP	2	*TTPA* (NM_000370)	c.553-1G>T	Homozygous	Ataxia with vitamin E deficiency (AVED)
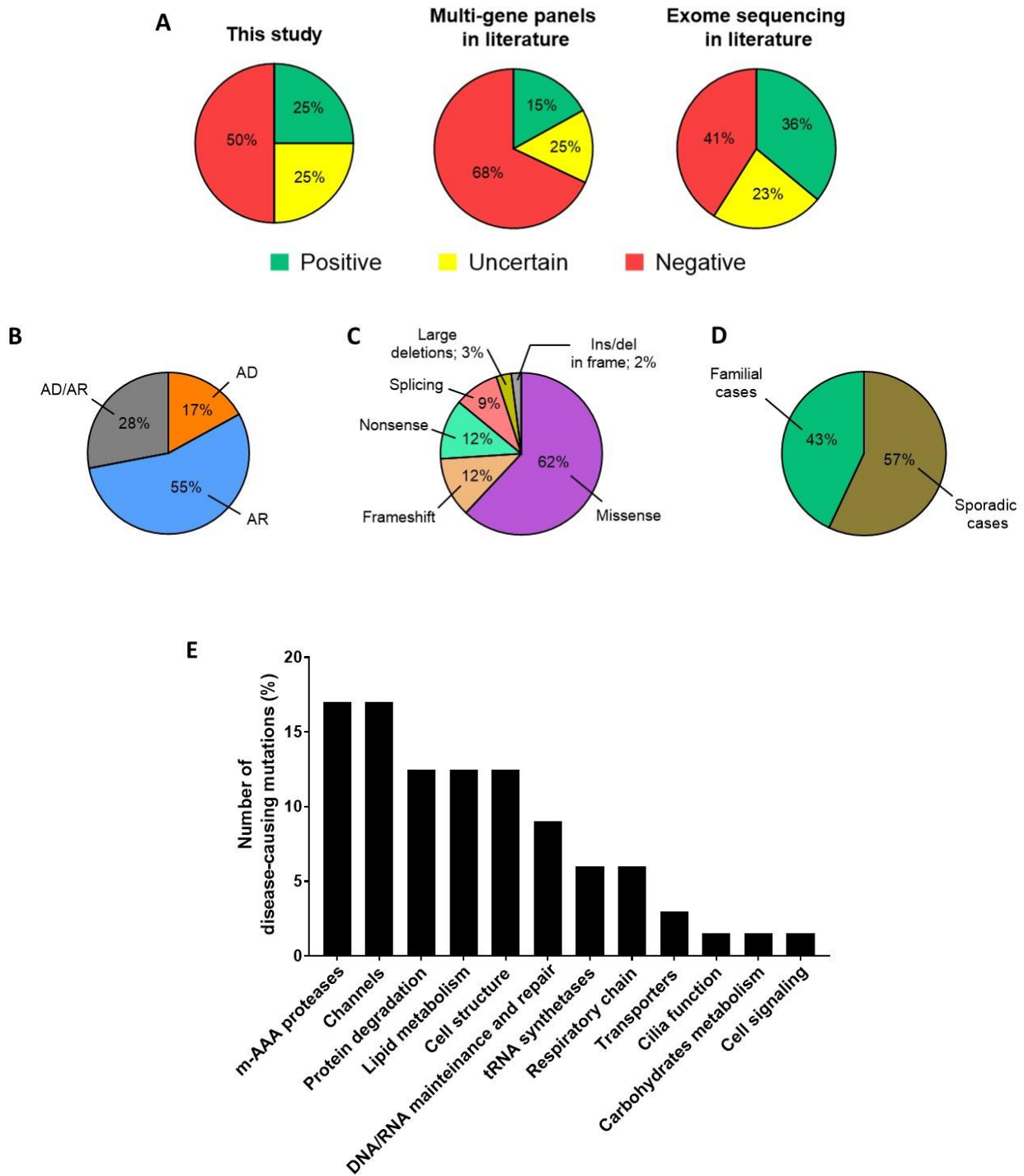
Figure 15. Summary of results of massive parallel sequencing. (A) Comparison of diagnostic rate between this study and average results in the literature. Our results displayed a higher percentage of positive cases compared to similar gene-panels studies in the literature, but the output is still lower than percentages reported in exome sequencing studies. The large number of patients with uncertain diagnoses suggests that segregation studies and detailed phenotypic characterization are required to reduce this uncertainty. (B) Frequency of inheritance patterns indicate a high involvement of recessive genes, and also a notable percentage of genes with both dominant and recessive patterns. (C) Frequency of different types of mutations denotes that two out of three variants are missense. (D) Frequency comparison between familial and sporadic cases among positive patients. (E) Molecular pathways of genes with disease-causing mutations among our cohort.

Despite representing only 17% of our whole cohort, familiar cases encompassed 43% (28/65) of all positive cases (Figure 15D).

SPG7 is the most common disease-causing gene in our cohort (9/65, 14%), followed by *STUB1* (7/65, 11%) and *PNPLA6* (5/65, 8%). Mutations in gene coding for m-AAA proteases (11/65, 17%) and channels (11/65, 17%) were the most common cause of disorder in our cohort (Figure 15E). Furthermore, we noticed a high frequency of mutations involved in metabolisms of complex lipids (8/65, 12.5%), in cell structure (8/65, 12.5%), and in DNA and RNA maintenance and repair (6/65, 9%) (Figure 15E). Nevertheless, we also found a high frequency in genes involved in protein degradation (8/65, 12.5%) (Figure 15E). We also identified 4 patients with mutations in tRNA-synthetases, 4 in genes encoding proteins implicated in the proper function of respiratory chain, and 2 in genes encoding membrane transporters (Figure 15E). Other mutations were founded in genes involved in cilia function, carbohydrates metabolism and cell signaling (1.5% each) (Figure 15E).

Application of exome sequencing in 12 highly selected patients with familial recurrence of early-onset pure cerebella ataxia, and negative to multi-gene panels applications, did not reveal clear-cut pathogenic variants. However, a number of VOUS are still under scrutiny (not shown).

High frequency of mutation in certain genes of our cohort (*SPG7, STUB1*) and pathogenic mutations described in genes associated with new or atypical phenotypes (namely, *AFG3L2, HARS, COQ4, COQ8A, RNF216, ATP13A2*) were selected among positive cases, limited to the availability of both biological samples beyond peripheral blood and/or enough detailed clinical data, and further investigated at the phenotypic and functional levels through different approaches.

4.2 *SPG7* results

4.2.1 Genetic outcomes

 Mutations in *SPG7* were the most common cause of HA in our cohort (9/64, 14%) (patients 43-51). Although NGS is not a valid tool to detect large insertion and deletions, we were able to identify a homozygous deletion of exon 2 in *SPG7* in 3 patients (patients 47, 48 and 50) because of complete lacking of coverage in that specific exon. Gene deletion was subsequently confirmed by MLPA analysis. Two patients harbored a truncating and a missense variant (patients 43-49), whereas we found two patients harboring the homozygous p.Ala510Val (patients 45 and 46) and two had the p.Ala510Val in compound heterozygosity with another missense mutation (patients 44 and 51).

4.2.2 Studies in skin fibroblasts

In our cohort, we could obtain a skin biopsy only from patient 45 (harboring the homozygous p.Ala510Val). Human fibroblasts of three other SPG7 patients not part of this study (SPG7 pt A, SPG7 pt B, SPG7 pt C), but available in our laboratory, were used to compare outcomes from the homozygous p.Ala510Val genotype with genotypes carrying at least one truncating mutation supposed to have higher differences when compared to controls.

Patient's 45 protein expression showed levels comparable to controls, in contrast to other SPG7 patients (Figure 16A and 16B). A similar trend was observed in oxygen consumption rate, where patient 45 did not show significant abnormalities (Figure 16D). On the other hand, its AFG3L2 expression tested by western blot resulted to be significatively reduced compared to controls, a feature shared by two other patients (Figure 16B and 16C).
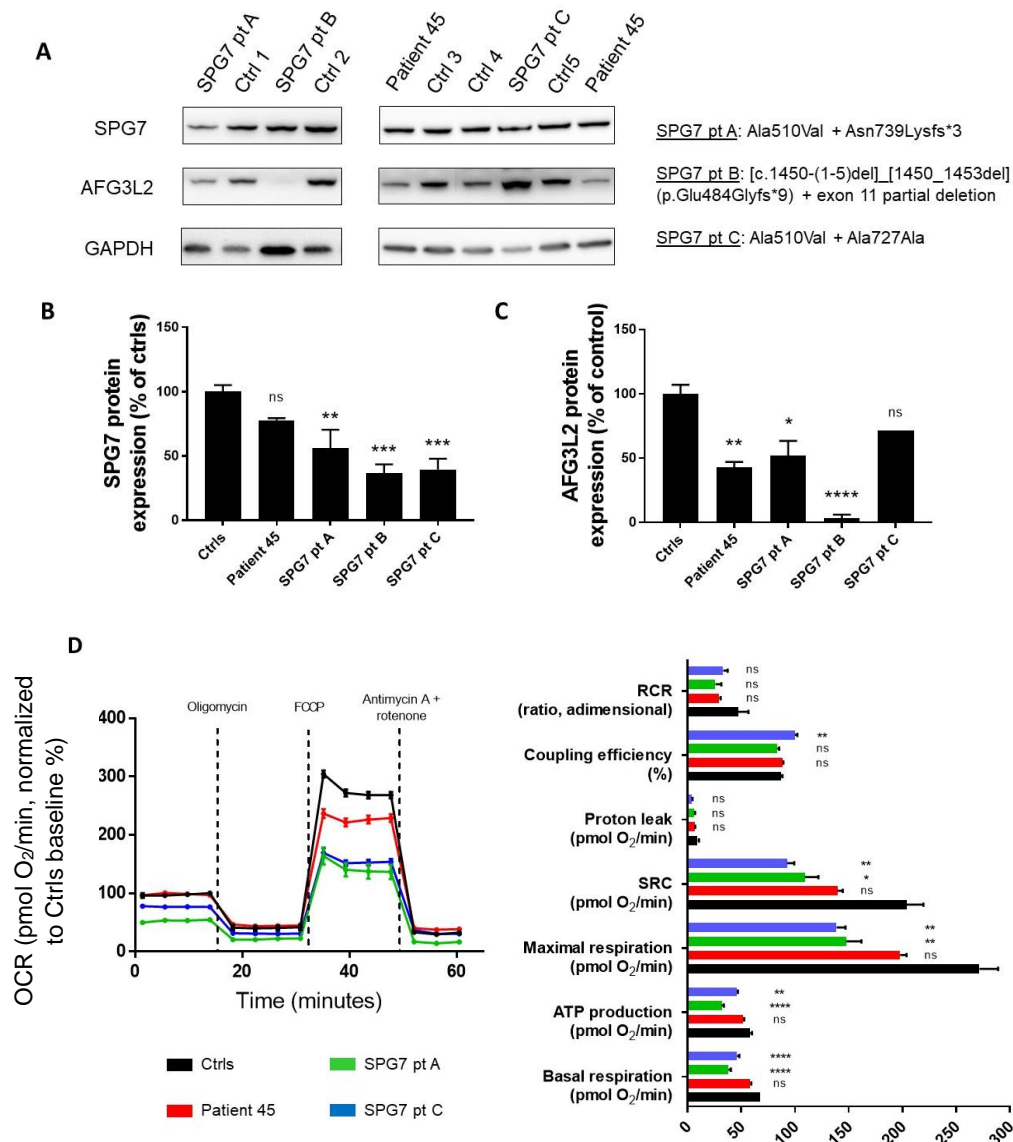

Figure 16. Legend in the next page

51

Figure 16. Functional studies in cultured skin fibroblasts from patient 45. (A) Representative Western blot used for detection of SPG7 and AFG3L2 proteins and their quantifications (B,C). (D) Despite slight differences in maximal respiration and spare respiratory capacity (SRC), all oxygen consumption rate (OCR) parameters in patient 45 were not statistically different when compared to controls, conversely to other patients harbouring truncating mutations and with a decrease paraplegin expression. Legend: SRC=spare respiratory capacity, RCR=cell respiratory control ratio.

4.2.3 Studies in *Drosophila melanogaster* knock-down models

Although our initial goal was to create a *SPG7* KO fruit fly model, tissue-specific downregulation can be helpful to understand which tissue is mainly affected upon paraplegin deficiency, an issue that cannot be questioned studying a model with a constitutional absence of *SPG7*. We took advantage of GAL4-UAS system to silence the *SPG7 Drosophila melanogaster* orthologue (*CG2658*, that we indicate as *dSPG7* for simplification) in a tissue-specific manner.

Two UAS lines expressing d*SPG7* RNAi were available (UAS-*dSPG7RNAi*, Bloomington Stock 55324; UAS-*dSPG7RNAi*, Bloomington Stock 57843). Both lines were tested by quantitative Real Time PCR, showing a downregulation of SPG7 expression by 90% and 10%, respectively (Figure 17A). Therefore, the UAS-*dSPG7RNAi 55324* was used for subsequent studies. We also observed that reduction of *dSPG7* expression did not influence the expression levels of *dAFG3L2* mRNA in flies (Figure 17B).

When downregulation was induced in the whole fly using *actin*-GAL4 line, a highly significant reduction of lifespan after one month was observed (Figure 17C). Indeed, median life span (MedLs, the reduction when 50% of the flies are dead) was 32 days, whereas maximum life span (MaxLs, reduction when all flies are dead) was 36 days. Since human SPG7 phenotype has a strong neuromuscular and neurodegenerative component, we tested locomotion with negative geotaxis assay, but no differences were noticed (Figure 17D). Paraplegin is known to be involved in several mitochondrial processes, included bioenergetics, and we tested aconitase (an enzyme that catalyzes the isomerization of citrate to isocitrate in Krebs cycle) activity and ATP production levels to assess damages in tricarboxylic acid cycle and electron chain transport, respectively. No differences were found in both assays after 1 week (Figure 17E and 17F), but a significative abnormal ATP production was detected after three weeks, suggesting a potential mitochondrial damage.
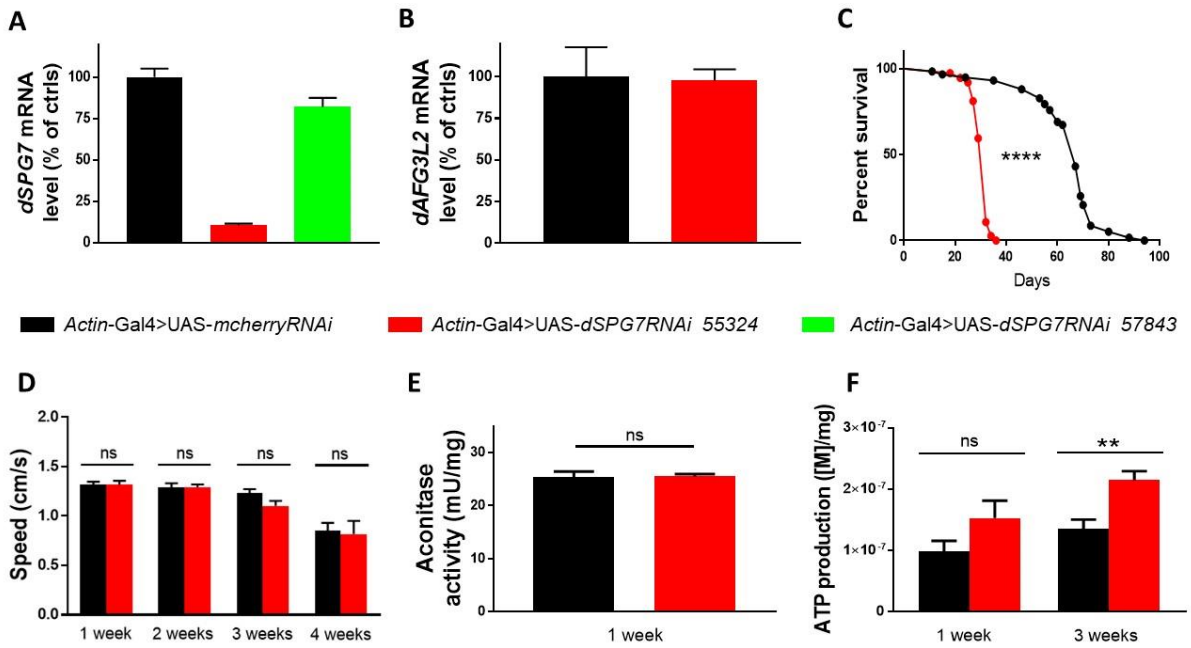
Figure 17. Effects of dSPG7 silencing in the whole flies. (A) Measurement of *dSPG7* and (B) *dAFG3L2* mRNA levels by quantitative real time PCR. (C) Lifespan assay. (D) Negative geotaxis assay. (E) Measurement of aconitase activity. (F) Measurement of ATP production.
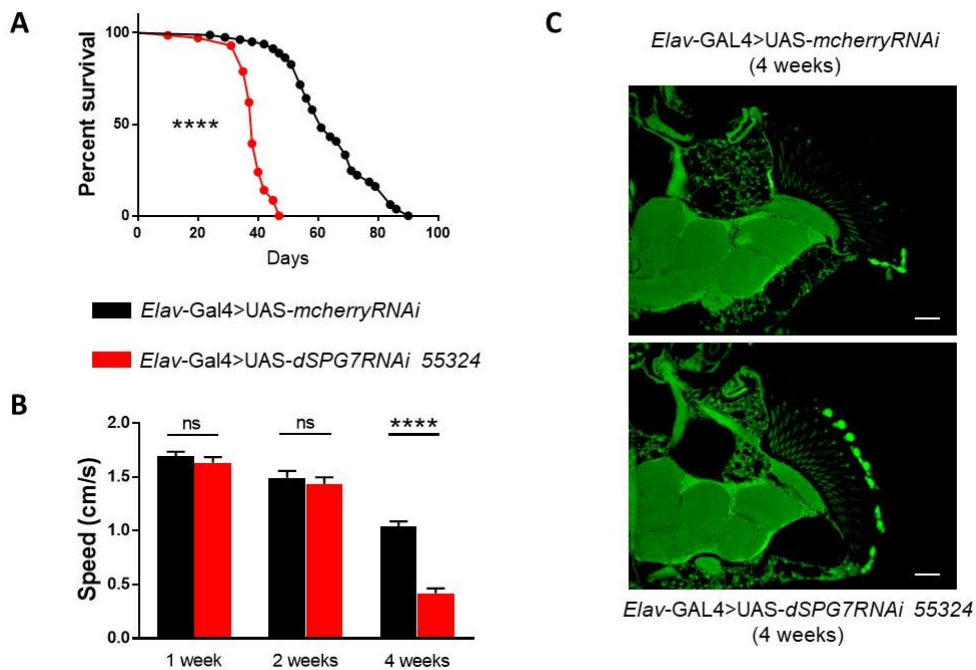
Figure 18. Effects of dSPG7 silencing in neurons. (A) Lifespan assay. (B) Negative geotaxis assay. (C) Brain sections (green signal comes from the autofluorescence caused by the eye pigment). Scale bars=100 µm.

When silencing is performed only in neurons using *Elav*-GAL4 line, we also observed a strong decrease in longevity after 40 days (MedLs 38 days, MaxLs 47 days, Figure 18A). Interestingly, we observed 50% reduction of locomotion after four weeks (Figure 18B). However, brain sections did not show vacuolization after 4 weeks, suggesting absence of neurodegenerative processes (Figure 18C).

When downregulation occurred only in muscles (a tissue enriched in mitochondria) using *Mef2-GAL4* line, we observed significant differences in lifespan (MedLs 43 days, MaxLs 79 days, Figure 19A),but not in locomotion (Figure 19B) and both muscle mitochondria and fibers morphology (Figure 19C).
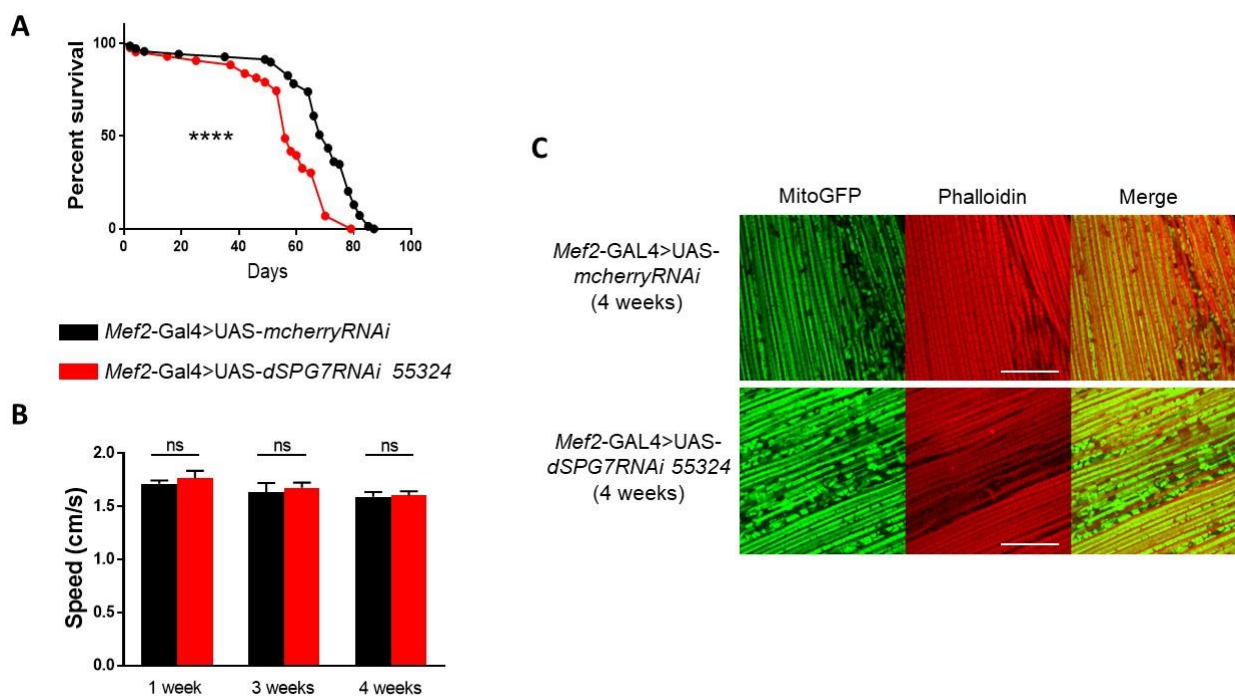
Figure 19. Effects of dSPG7 silencing in muscles. (A) Lifespan assay. (B) Negative geotaxis assay. (C) Indirect flight muscles sections. Green signal in left panels indicates GFP labelling mitochondria, red signal in central panels indicates phalloidin staining F-actin. Scale bars=25 µm.

To check if nervous system defects are attributable to glia cell rather than neurons, we downregulated *dSPG7* in glia using *Repo*-GAL4 line but did not detect differences in locomotion (Figure 20A).

Considering that the human clinical phenotype is characterized by significant cerebellar involvement, we immunolabeled the mushroom body (the cerebellum equivalent in fruit fly) in flies where paraplegin was silenced in that structure (*MB*-GAL4 line) and observed a moderate, signal decrease (Figure 20B).

Also, we tested the response to induced oxidative stress but this condition did not impact on lifespan (MedLs 8 days, MaxLs 10 days, figure 20C) and brain integrity (Figure 20D).

Finally, *dSPG7* overexpression in the whole fly (3X normal) (Figure 21A), did not provide differences to oxidative stimuli (MedLs 9 days, MaxLs 9 days, Figure 21B) or in locomotion in normal conditions (Figure 21C).
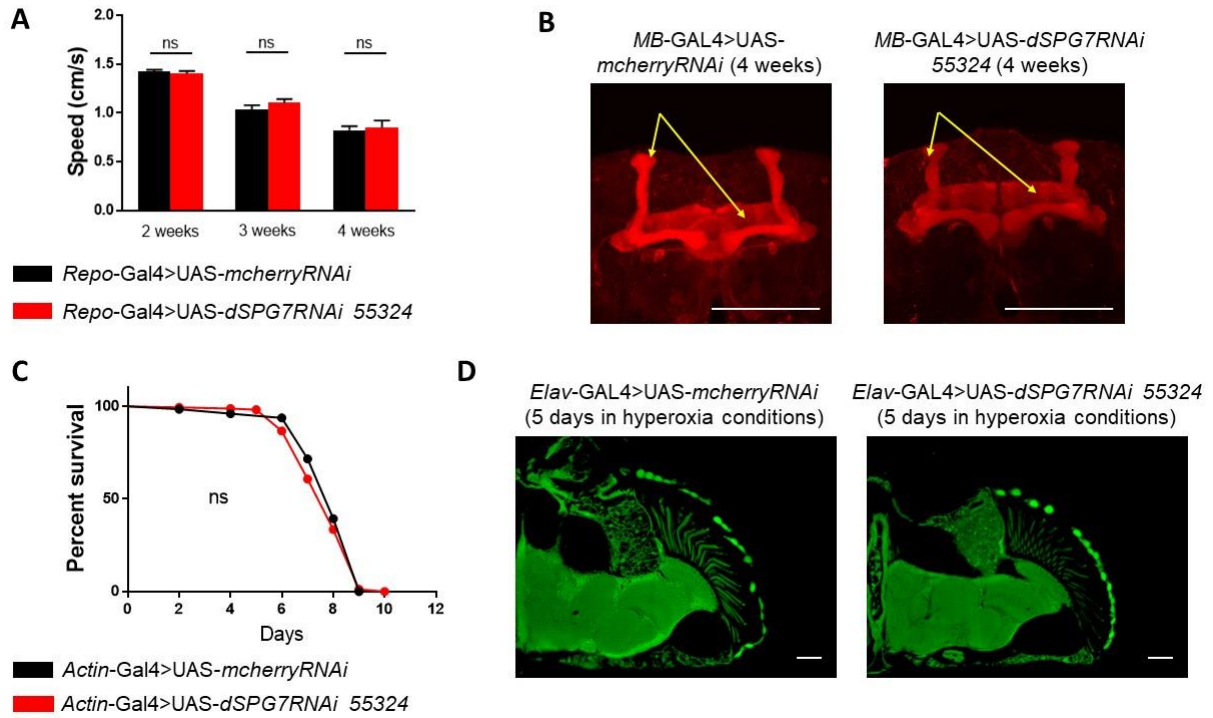
Figure 20. Effects of dSPG7 silencing in other cell populations or under oxidative stress. (A) Negative geotaxis assay in *dSPG7*-silenced glia cells. (B) Immunostaining of the *dSPG7*-downregulated mushroom bodies (indicated by yellow arrows, red signal indicates FasII antibody staining). (C) Lifespan assay in hyperoxia conditions when *dSPG7* is silenced in the whole flies. (D) Brain sections in similar conditions when *dSPG7* occurs only in neurons. Scale bars=100 µm.
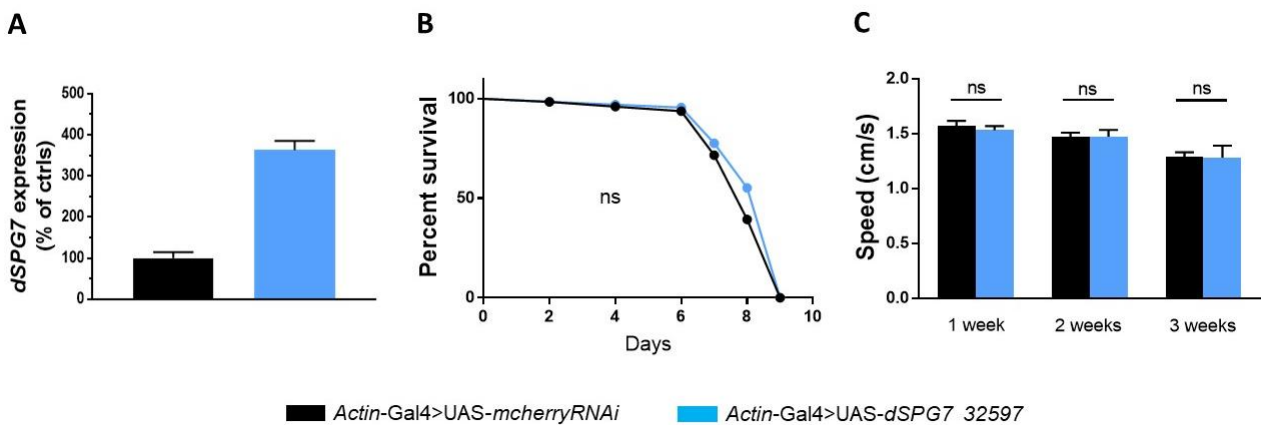
Figure 21. Effects of dSPG7 overexpression in the whole flies. (A) Measurement of *dSPG7* mRNA levels by quantitative real time PCR. (B) Lifespan under hyperoxia conditions. (C) Negative geotaxis assay.

4.2.4 Generation of a *Drosophila melanogaster SPG7* knock-out model to set up a novel *SPG7^{Ala510Val}* KI model

The p.Ala510Val mutation is the most common variant in SPG7 patients and it is tempting to model its effects *in vivo*. Having tested that alanine 510 is highly conserved in different species including fruit fly (Figure 22A), we set modeling in Drosophila. To do this, we first created a *SPG7* knock-out model. Creation of a knock-out by precise insertion, through homologous recombination, of a large cassette containing a fluorescent marker permits to screen successful gene editing events among flies were recombination does not occur (step 1 to step 2 in figure 22B). A new exchange between cassette previously inserted with the gene harboring the mutation of interest causes the loss of the fluorescent marker, giving the possibility to discriminate again where recombination arises (step 2 to step 3, in figure 22B). Therefore, creation of a KO model is critical because the alternative option of a direct exchange with the sequence bearing the mutation of interest would have not provided (at least with the method we used) markers to distinguish where recombination events occur.

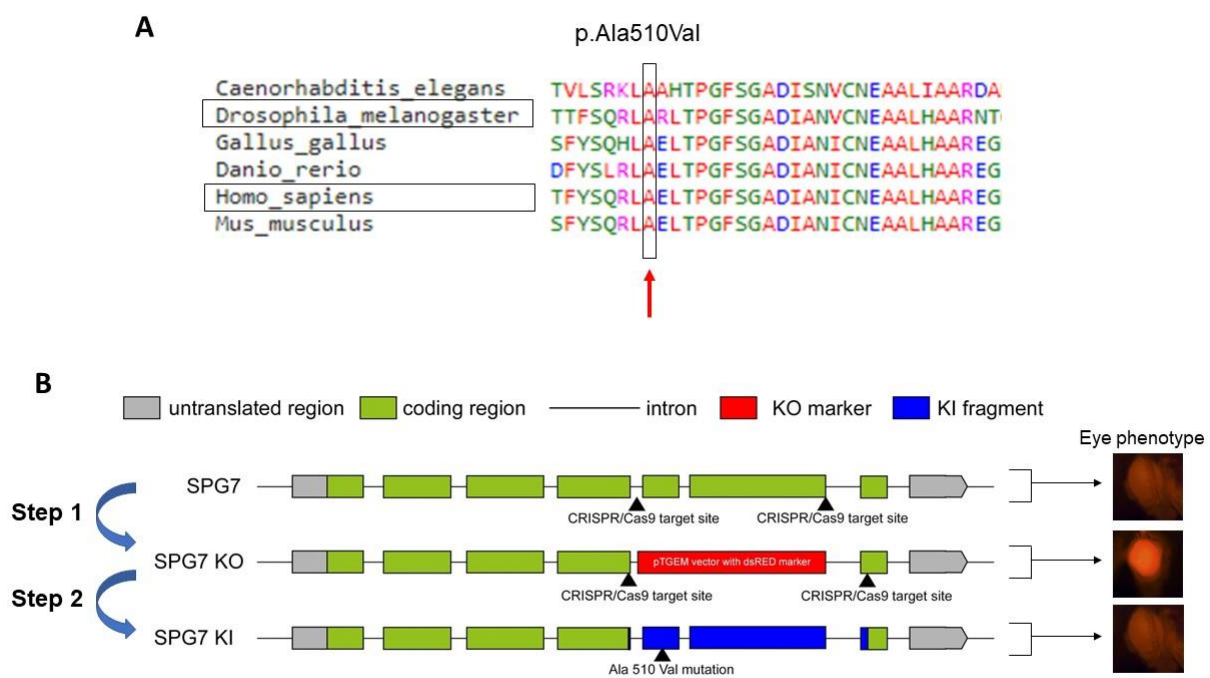*Figure 22. Strategy for generation of the p.Ala510Val SPG7 knock-in Drosophila model.* (A) Phylogenetic analysis of Ala510Val among different species. (B) Generation of a knock-in model in fruit fly using a two-step strategy through homologous recombination Initial gain and subsequent loss of red fluorescent marker in eyes (right) allows to recognize successful recombination events.

To generate the *dSPG7* KO strain, 700 embryos were injected with a solution containing both vectors carrying the gRNAs and pT-GEM/d*SPG7*hom1/d*SPG7*hom2 vector (see Material and methods for details). Around 140 of them survived until adulthood, they were crossed to white mutant flies and the offspring of those crosses was screened to identify flies containing the red fluorescent marker 3xP3 in the eye. Positive transformants were identified in 10 of the crosses. Genotyping of these transformant lines by PCR showed correct and precise substitution of the gene as expected (Figure 23A). Similar to previous work (Pareek et al., 2018), we observed a severe death rate (KO line 3 MedLs 37 days and MaxLs 43 days, KO line 4 MedLs 32 days and MaxLs 41 days, Figure 23B), and impaired locomotion (decrease of 90% after four weeks. Figure 23C). Moreover, susceptibility to oxidative stress and defects in electron transport chain complexes were confirmed by a significative reduction of survivals in hyperoxia conditions (MedLs 6 day, MaxLs 8 days, Figure 23D) and an abnormal ATP production (Figure 23E), respectively. At the time of this writing, we are going to inject the vector harboring p.Ala510Val mutation (see Materials and methods) in *dSPG7* KO flies to create the humanized knock-in model, which represents our main purpose. Nevertheless, this model might be useful also for *in vivo* validation of human variants.
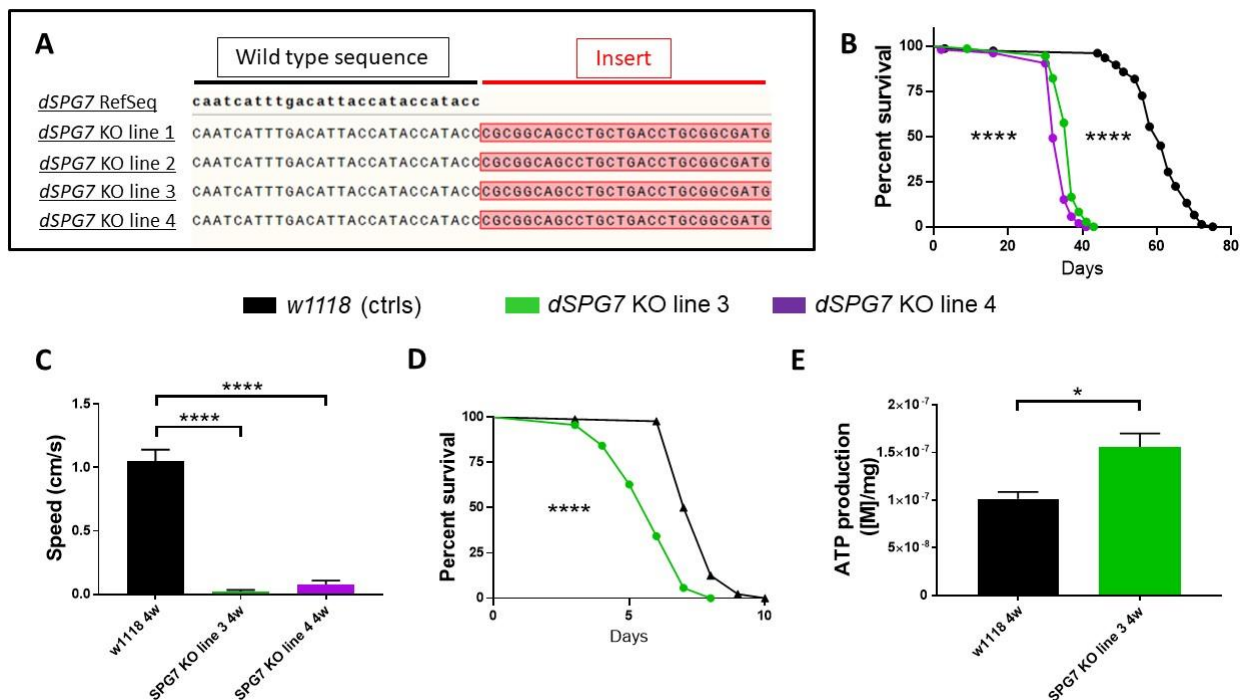
Figure 23. Generation and characterization of SPG7 KO Drosophila. (A) Genotyping of transformant *SPG7* KO lines demonstrates the successful inclusion of the insert (labelled in red). Only four of the ten lines are shown. (B) Lifespan assay. (C) Negative geotaxis assay. (D) Lifespan assay under hyperoxia conditions. (E) Measurement of ATP production. Only one or two lines were used for functional characterization of the knock-out model since our aim was only to confirm and compare the resulting phenotype with the one obtained by Pareek and collaborators (Pareek et al., 2018).

4.3 *AFG3L2* results

4.3.1 Genetic outcomes

 We identified pathogenetic mutations in *AFG3L2* in two patients in our cohort. Patient 3 harboured p.Arg702Gln causing SCA28, a rare variant (MAF=3.98×10^{-6}) whose pathogenicity was already described in literature and found in a family with dominant history, showing moderate cerebellar atrophy and variable penetrance and expressivity (Di Bella et al., 2010). Patient 2 was a child characterized by episodic ataxia, and harbored a frameshift p.Val212Glyfs*4 in compound heterozygosity with a rare missense p.Val723Met (MAF=1.79×10^{-4}) recently described as disease-causing by others (Tunc et al., 2019).

4.3.2 Studies in skin fibroblasts

 In patient 2 we observed a significantly reduced AFG3L2 protein levels (~50%, Figure 24A), with normal levels of paraplegin (Figure 24B).

Figure 24. Protein expression analysis in patient 2. Representative Western blot of AFG3L2 (A) and SPG7 (B) in patient 2.

4.3.3 Studies in *Drosophila melanogaster* knock-down models

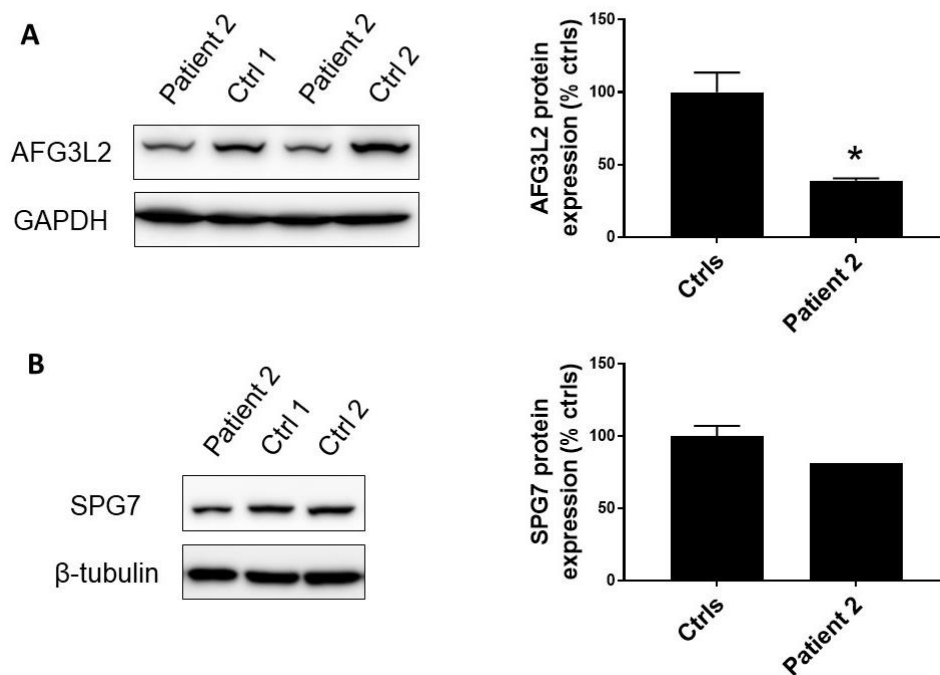AFG3L2 is a m-AAA that forms homo-exameric complex, and it has a critical role to form hetero-hexameric complexes with paraplegin, hence deserving modeling in fruit fly. As well as for *SPG7*, we used GAL4-UAS system to silence the *AFG3L2 Drosophila melanogaster* orthologue (*CG6512*, that we indicate as *dAFG3L2* for simplification) in a tissue-specific manner.

One UAS line expressing *dAFG3L2* RNAi was initially available (UAS-*dAFG3L2RNAi*, Bloomington Stock 50524), and qPCR showed a downregulation of 70% of AFG3L2 mRNA expression (Figure 25A), whereas no abnormalities in *dSPG7* mRNA levels were noticed (Figure 25A). Downregulation in the whole fly using *actin*-GAL4 triggered some developmental lethality but escapers were detected, though they died within the first month (MedLs 20 days, MaxLs 32 days, Figure 25B), and all had a strong lethality in hyperoxia conditions (MedLs 4 days, MaxLs 7 days Figure 25C). Locomotion resulted to be highly reduced (Figure 25D), with a 30% reduction after both one and two weeks, and a more significant one (75%) after three weeks. No living flies were available at 4 weeks to investigate this time-increasing decrease of locomotion (see Figure 25B). Determination of aconitase enzyme activity was significantly reduced after one week (25%, figure 25E), whilst ATP production seemed to show a tendency to an abnormal increase but without statistically significant differences (Figure 25F).
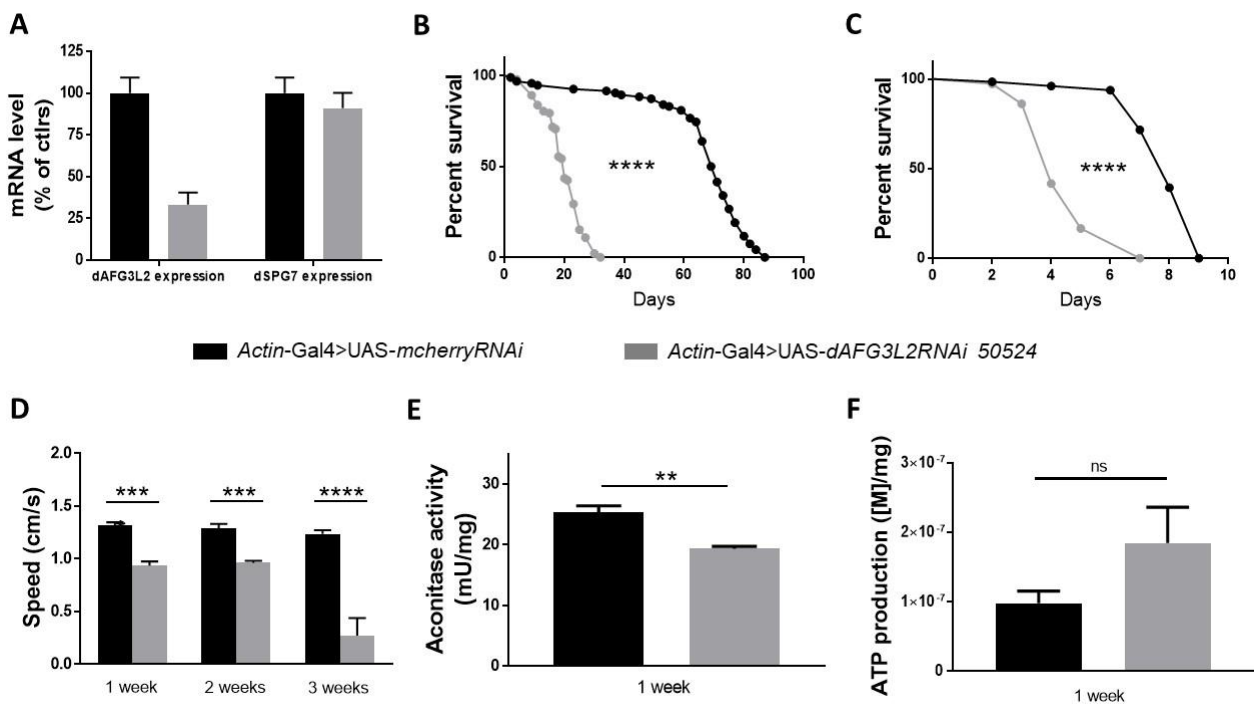

Figure 25. Effects of dAFG3L2 silencing in the whole flies. (A) Measurement of *dAFG3L2* and *dSPG7* mRNA levels by quantitative real time PCR. (B) Lifespan assay. (C) Lifespan assay in hyperoxia condition. (D) Negative geotaxis assay. (E) Measurement of aconitase activity. (F) Measurement of ATP production.

Restricted downregulation to muscles using *Mef2*-GAL4 showed early death (MedLs 5 days, MaxLs 9 days, Figure 26A) and a severe reduction of locomotion (Figure 26B), accompanied by a significative decrease of ATP production (75%, Figure 26C). Moreover, muscle histology showed abnormal mitochondrial aggregation and thinner muscle fibres. (Figure 26D). All these data suggest a strong involvement of *dAFG3L2* in muscle function caused by mitochondrial impairments.

Even more severe effects were remarked when downregulation occurred in neurons using *Elav*-GAL4, showing a developmental delay in terms of days achieved to hatch from pupa stages compared to controls, and a sudden death within 1-2 days. Nonetheless, no differences were observed in brain histology performed in flies hatched from few hours compared to controls (Figure 27A). Bypassing the developmental downregulation of *dAFG3L2* in neurons using TubulinGAL80^{ts} tool reverted differences in survival rate (MedLs 28 days, MaxLs 33 days Figure 27B) and locomotion (Figure 27C), whereas brain vacuolization was observed after four weeks (Figure 27D).

In addition, no differences were noticed in locomotion when downregulation occurs in glia (Figure 28A), whereas a slight decrease of given by FasII signal (that labels mushroom bodies) was seen when *dAFG3L2* was silenced in mushroom bodies (Figure 28B). All these findings suggest that *AFG3L2* has a pivotal role in neuronal development, at least in fruit flies.
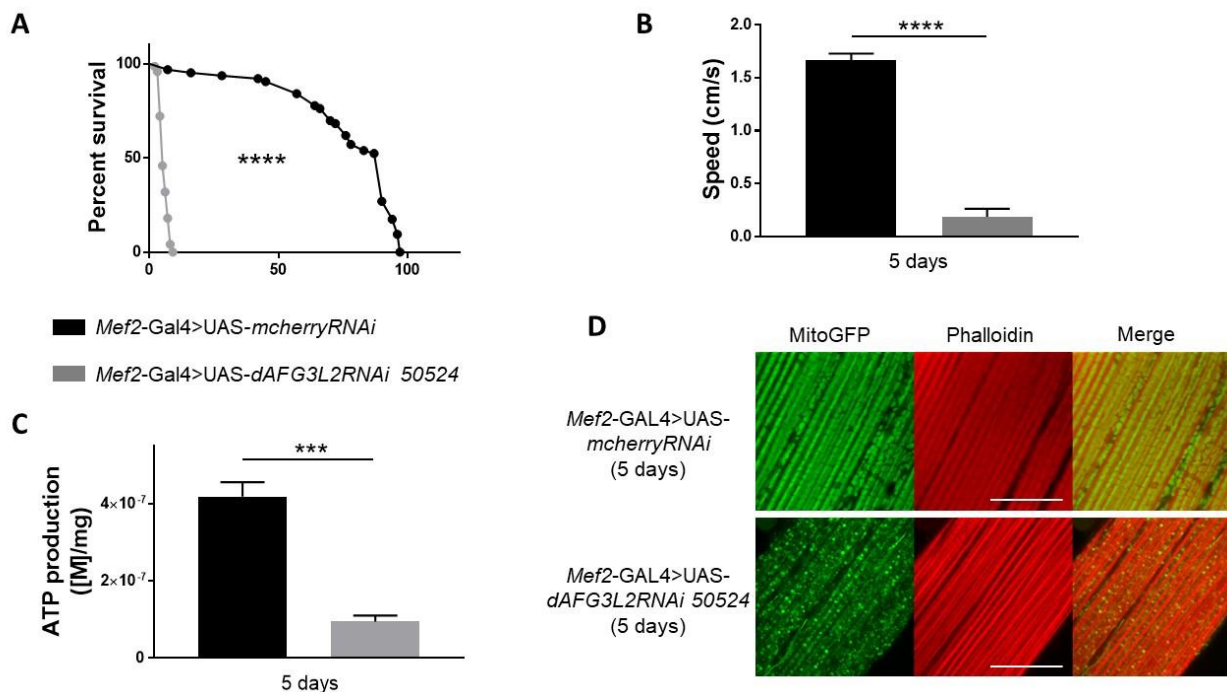
Figure 26. Effects of dAFG3L2 silencing in muscles. (A) Lifespan assay. (B) Negative geotaxis assay. (C) Measurement of ATP production. (D) Indirect flight muscles sections. Green signal in left panels indicates GFP labelling mitochondria, red signal in central panels indicates phalloidin staining F-actin. Scale bars=25 µm.
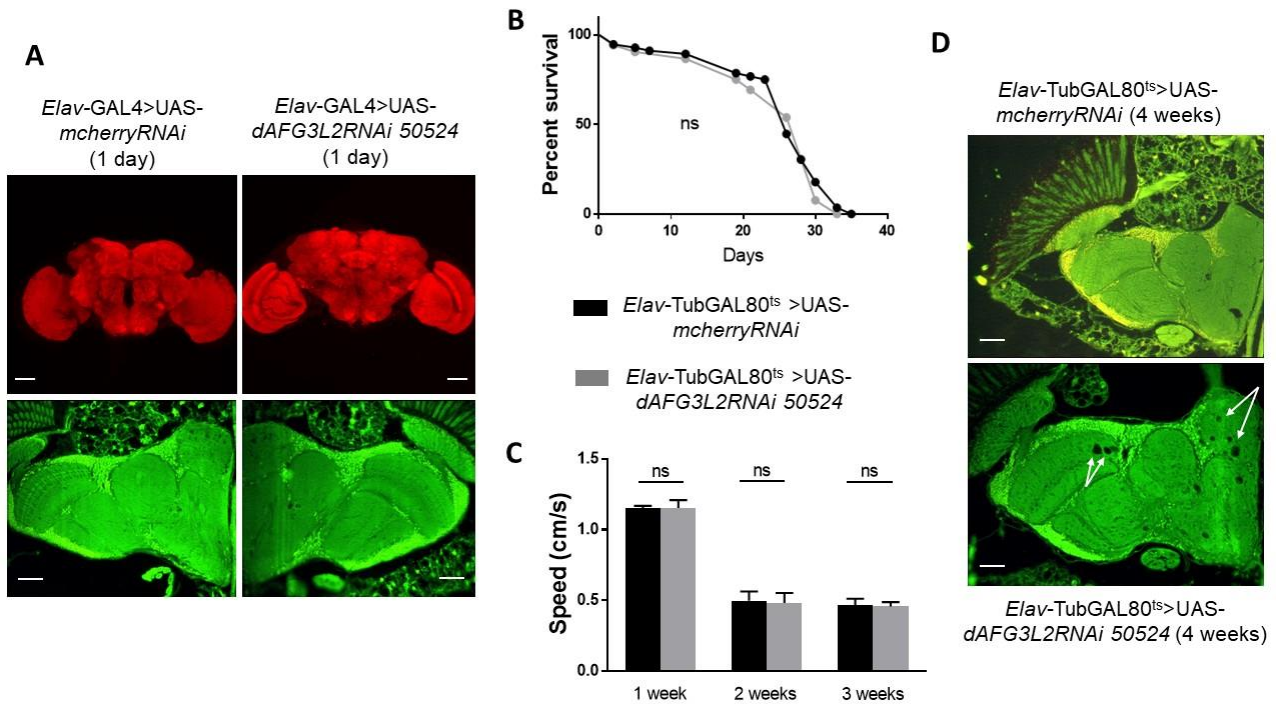
Figure 27. Effects of dAFG3L2 silencing in neurons. (A) Histological studies in flies' brain. Upper panels show immunostaining of whole brain (red signal indicates Bruchpilot antibody that labels all neurons), lower panels illustrate brain section (green signal comes from the autofluorescence due to eye pigment). (B) Lifespan assay. (C) Negative geotaxis assay. (D) Brain sections, white arrows indicate vacuoles. Scale bar=50 μm.
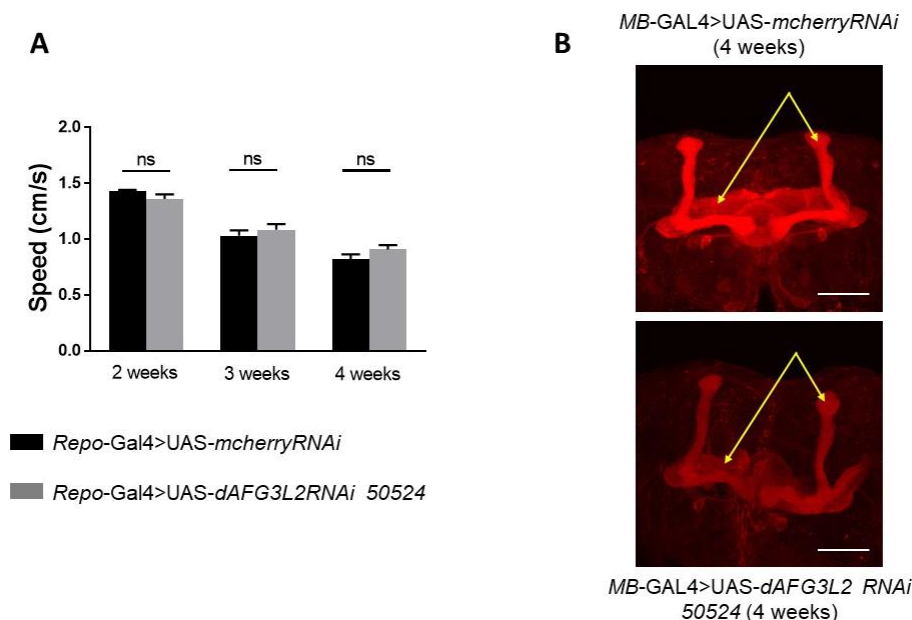
Figure 28. Effects of dAFG3L2 silencing in glia cells and mushroom bodies. (A) Negative geotaxis assay when *dAFG3L2* downregulation occurs in glia cells. (B) Immunostaining of the *dAFG3L2*-downregulated mushroom bodies (indicated by yellow arrows, red signal indicates FasII antibody staining). Scale bar=50 μm.

4.4 *HARS* results

4.4.1 Genetic studies

We detected compound heterozygous mutations in *HARS* in two unrelated patients in our cohort. Patient 16 is a child presenting spastic ataxia, dysmetria, progressive dysarthria, mild intellectual disability, movements disorders and cerebellar atrophy. He harbors a frameshift p.Val244Cysfs*6 and a missense p.Asp206Tyr mutation. Patient 17 is a woman presenting congenital ataxia, mild intellectual disability, and dystonic postures. She carries an in-frame insertion p.Leu305dup and a missense p.Ile465Leu. The similarly affected younger sister harbors the same mutations

The p.Val244Cysfs*6 in patient 16 occurs on the paternal allele whereas p.Asp206Tyr was inherited from the mother (Figure 29A). The p.Leu305dup in patient 17 was on the maternal allele whereas we inferred that p.Ile465Leu was inherited from the father, though we could not genotype paternal DNA (Figure 29B). Whilst the p.Ile465Leu has a very low frequency in gnomAD (MAF=1.22×10^{-5}), the other variants are absent in publicly accessible databases and in our in-house multi-gene panel and WES dataset. Most *in silico* predictions score a negative effect for these mutations on protein function. Of note, p.Asp206Tyr and p.Leu305dup lay in the catalytic domain of HARS, critical for enzymatic activity, whereas p.Ile465Leu is located in the tRNA binding domain, important for recognizing the correct tRNA molecules (Figure 30A).
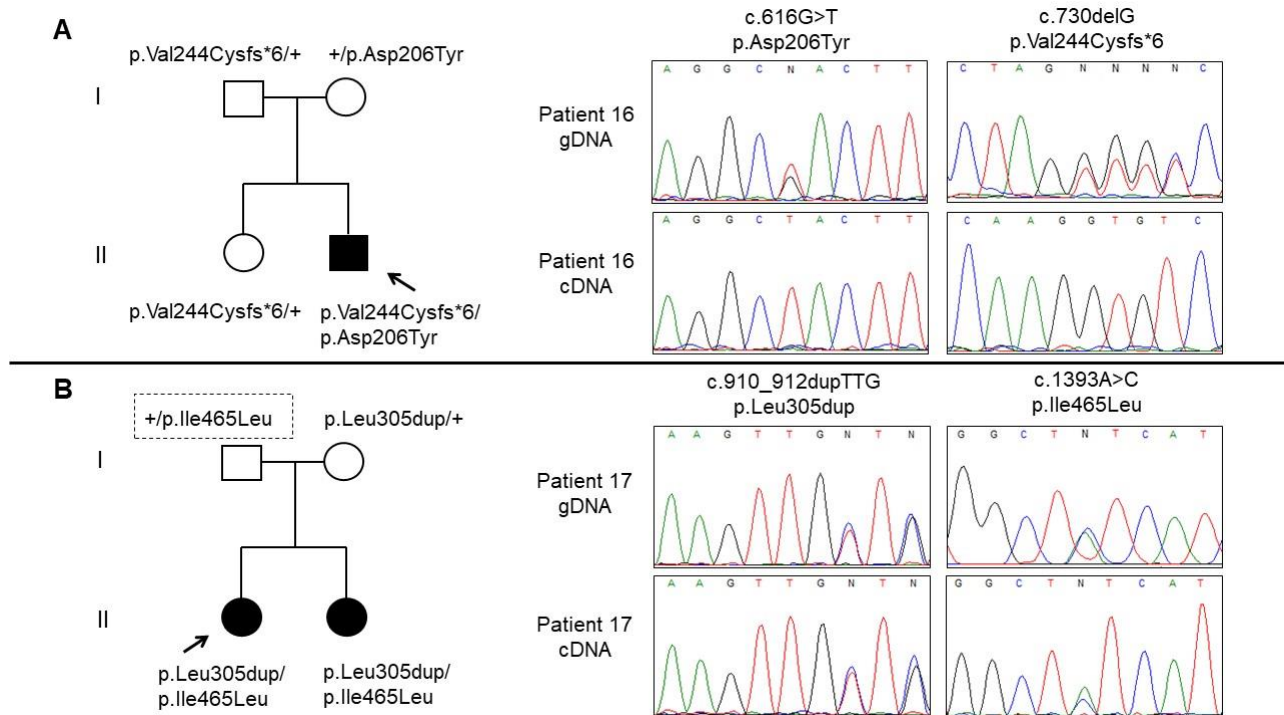
Figure 29. Legend in the next page.

Figure 29. Pedigrees, HARS variants confirmations and cDNA mutation analysis. The family pedigree of Patient 16 (A) and Patient 17 (B) shows that all healthy relatives carried a single heterozygous mutation. cDNA analysis demonstrates that that only mRNA bearing p.Asp206Tyr was expressed in Patient 16 (A), whereas both mRNA produced were stable in Patient 17 (B). Dotted box indicates the inferred paternal genotype (from Galatolo et al., submitted).
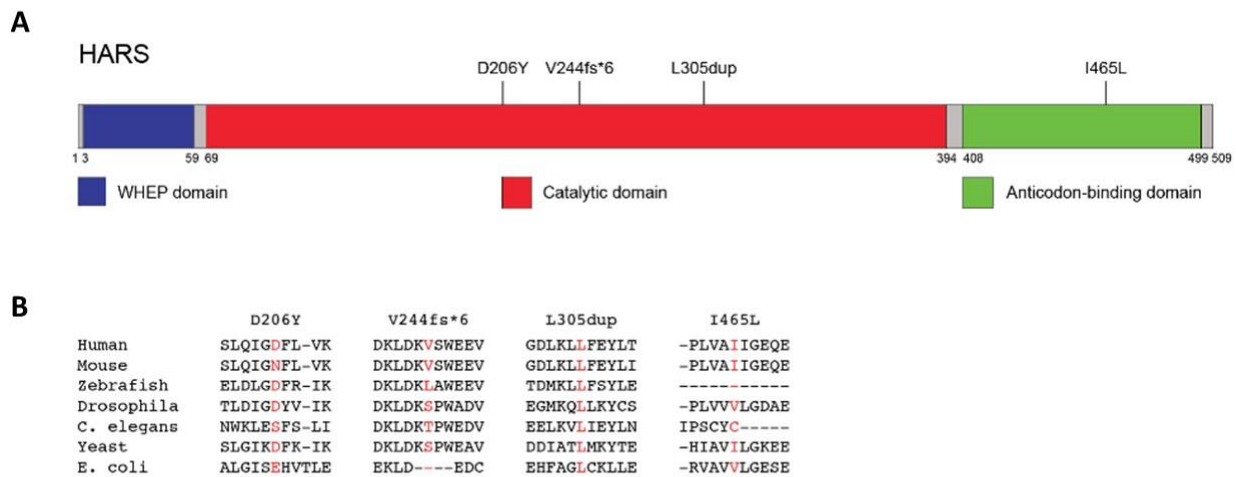
*Figure 30. Mutations in HARS. Disease-causing variants identified in patients 16 and 17. (*A) Scheme of HARS protein (UniProt P12081) showing the localizations of pathogenic mutations found. (B) Phylogenetic conservation of *HARS* variants identified in this study (from Galatolo et al., submitted).

4.4.2 Studies in skin fibroblasts

To determine the effect of each *HARS* genotype on mRNA and protein levels, cDNAs and whole cell lysates were obtained from primary fibroblasts isolated from both patients as well as from the affected sister of patient 17, and real-time qPCR and Western blot analysis were performed. In patient 16, the p.Asp206Tyr mutation but not the Val244Cysfs*6 was expressed at the cDNA level suggesting that the mRNA carrying the frameshift was unstable or subjected to nonsense-mediated decay (Figure 29A). In contrast, both variants in the two sisters were expressed at the mRNA level (Figure 29B). In patient skin fibroblasts, quantitative PCR and Western blotting indicated a significant reduction of HARS mRNA (Figure 31A) and protein expression (Figure 31B), respectively. Moreover, aminoacylation activity showed a dramatic reduction of tRNA charging in all patients (Figure 31C).
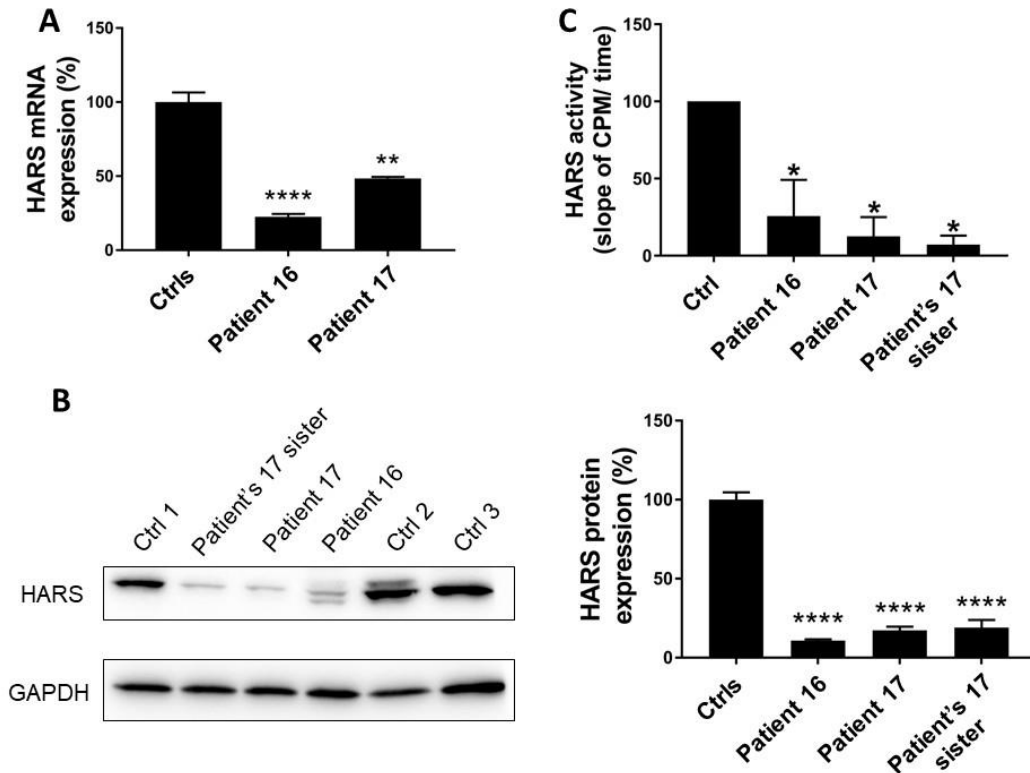
Figure 31. HARS mutations. Functional studies in cultured skin fibroblasts from patients 16 and 17. (A) Determination of *HARS* mRNA levels through real time quantitative PCR. (B) Representative western blot (left) used for HARS protein detection, and protein level quantification (right). (C) Measurement of HARS aminoacylation activity. * indicates $p<0.02$ in panel C (from Galatolo et al., submitted).

4.4.3 Studies in *Saccharomyces cerevisiae*

 To establish the functional consequences of each allele *in vivo*, yeast complementation assays were employed. These analyses revealed that: (1) wild-type human *HARS* complement yeast cell growth; (2) both p.Asp206Tyr and p.Ile465Leu alleles allow growth rates similar to wild-type, whilst (3) p.Val244Cysfs*6 and p.Leu305dup alleles show data consistent with a loss of function effect (Figure 32).

Figure 32. Yeast complementation assay for HARS mutations. Yeast lacking endogenous *HARS* were transformed with a vector with no *HARS* insert ('empty'), or with vectors containing wild-type or variant *HARS*. Cultures were plated undiluted or diluted (1:10 or 1:100) on media containing 5-FOA and grown at 30°C. Two independently transformed colonies were tested (upper and lower panels) (from Galatolo et al., submitted).

4.5 *COQ4* results

4.5.1 Genetic outcomes

We detected biallelic mutations in *COQ4* in two patients in our cohort. Patient 10 showed the c.577C>T, p.Pro193Ser in compound heterozygosity with the c.718C>T, p.Arg240Cys, and patient 11 harbored the c.284G>A, p.Gly95Asp and c.305G>A, p.Arg102His variants. Patient 10 is a 4-years-old boy showing early ataxia and developmental disorders and presents both normal CoQ_{10} levels and oxidative phosphorylation activities of mitochondrial respiratory chain complexes in skeletal muscle biopsy. Patient 11 is a 19-year-old girl presenting epilepsy and developmental delay since early childhood and developing a progressive spastic ataxia disorder associated with upper cerebellar vermis atrophy at brain imaging. In both families, mutations segregated with the disease (Figure 33A and 33B). In silico, all variants were scored as damaging and phylogenetically conserved (Figure 34A), and p.Arg102His and p.Pro193Ser were absent in public and in-house

databases, whereas p.Gly95Asp and p.Arg240Cys (already described before in Brea-Calvo et al. 2015, and Chung et al. 2015) were rare (MAF=4.01×10^{-6} and 1.90×10^{-4}, respectively).
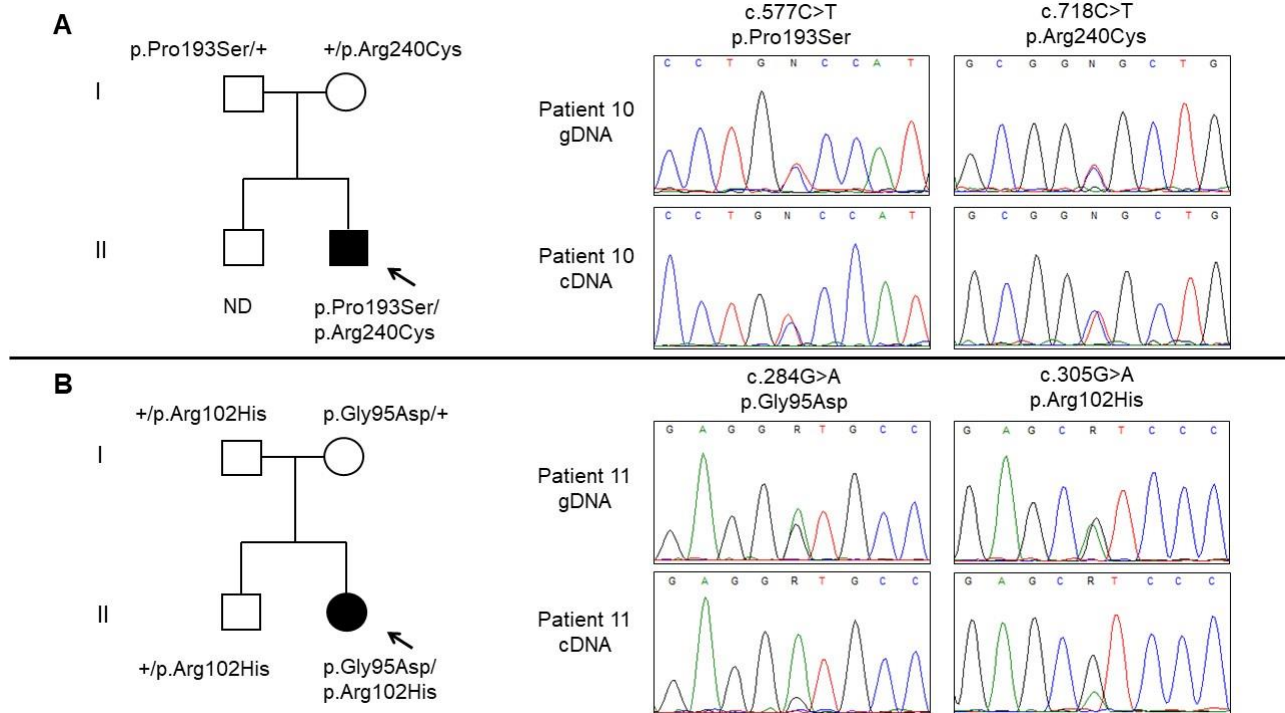
Figure 33. Pedigrees, COQ4 variants confirmations and cDNA mutation analysis.

4.5.2 Studies in skin fibroblasts

 To determine the effect of *COQ4* mRNA and protein levels, cDNAs and whole cell lysates were obtained from primary fibroblasts isolated from patients. cDNA analysis showed that both mRNA produced were stable in patient 10 (Figure 33A), whereas mRNA harboring p.Arg102His seemed present but under-expressed in patient 11 (Figure 33B). Both quantitative real time PCR (Figure 34B) and Western Blot (Figure 34C) showed no differences compared to controls. Patient 10 displayed an abnormal oxygen consumption rate (OCR) hence suggesting the presence of mitochondrial dysfunctions, whereas patient 11 OCR was similar to controls (Figure 34D).
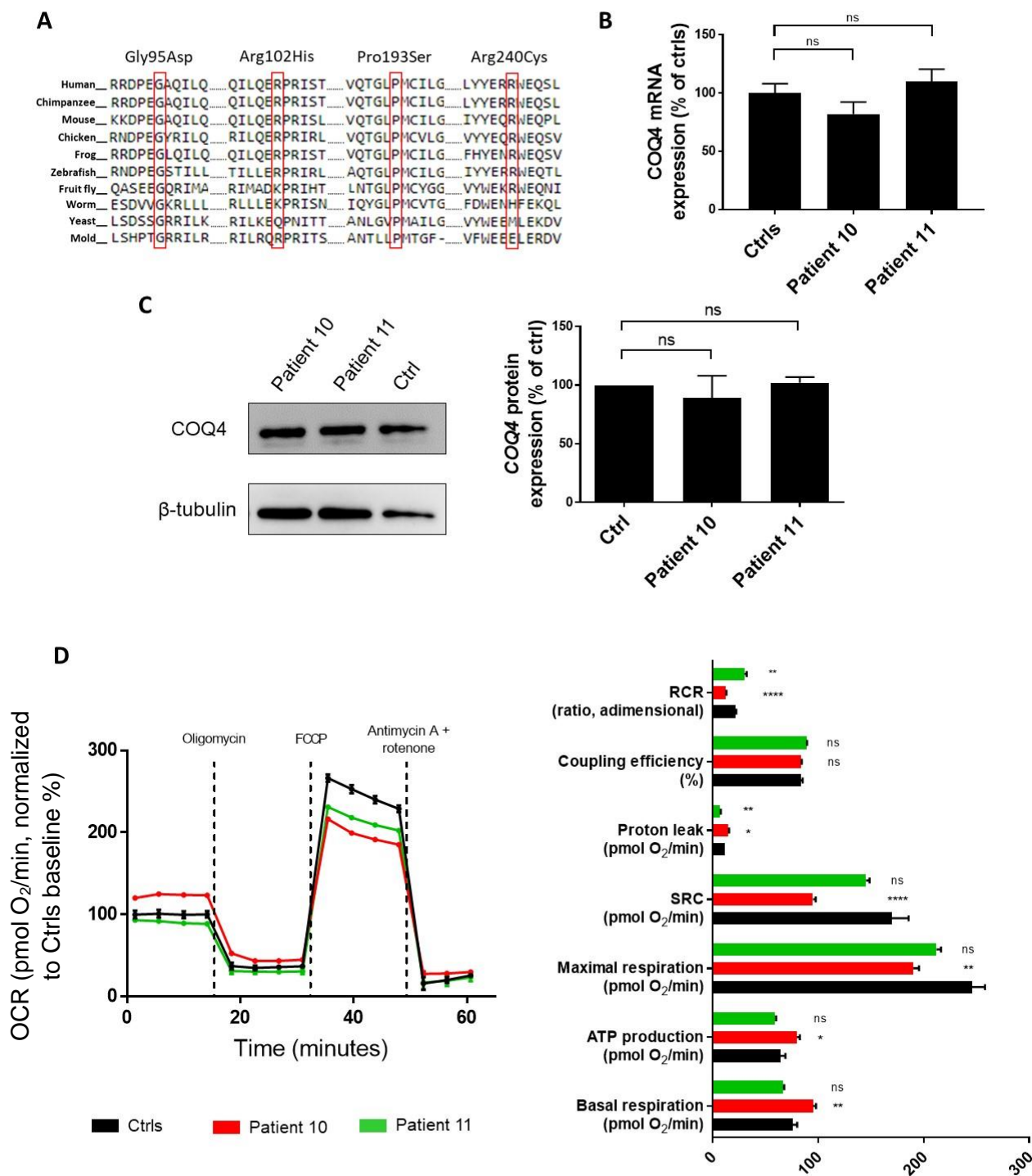
Figure 34. Phylogenetic analysis of COQ4 mutations and functional studies in cultured skin fibroblasts from patients 10 and 11. (A) Conservation of *COQ4* variants among different species. (B) Measurement of *COQ4* mRNA levels by real time quantitative PCR. (C) Representative western blot used for COQ4 protein detection (left) and protein level quantification (right). (D) Measurement and analysis of oxygen consumption rate (OCR). Legend: SRC=spare respiratory capacity, RCR=cell respiratory control ratio.

4.5.3 Studies in *Saccharomyces cerevisiae*

When tested by yeast complementation assay (Figure 35), p.Gly95Asp, p.Pro193Ser and p.Arg240Cys showed a complete absence of growth, indicating loss-of-function effects. p.Arg102His exhibited a hypomorphic effect instead (Figure 35).
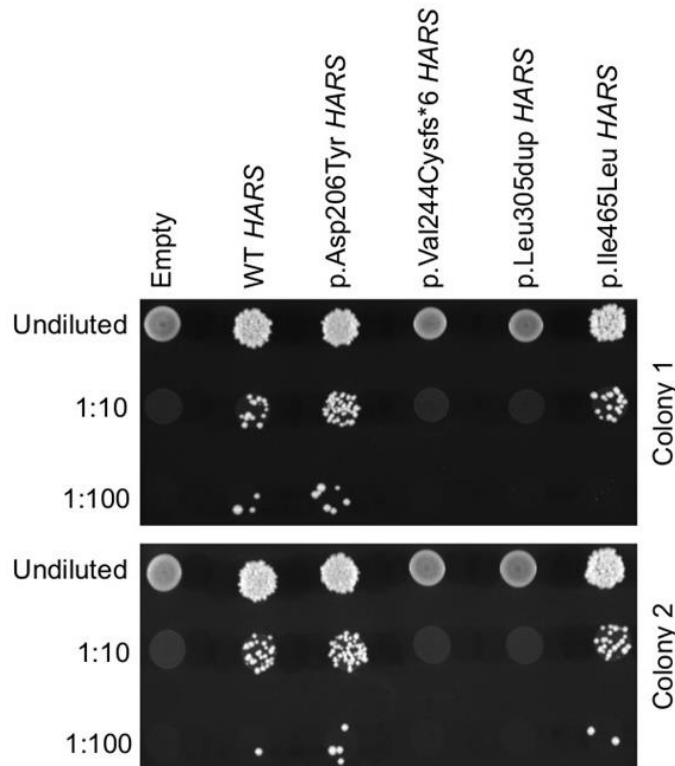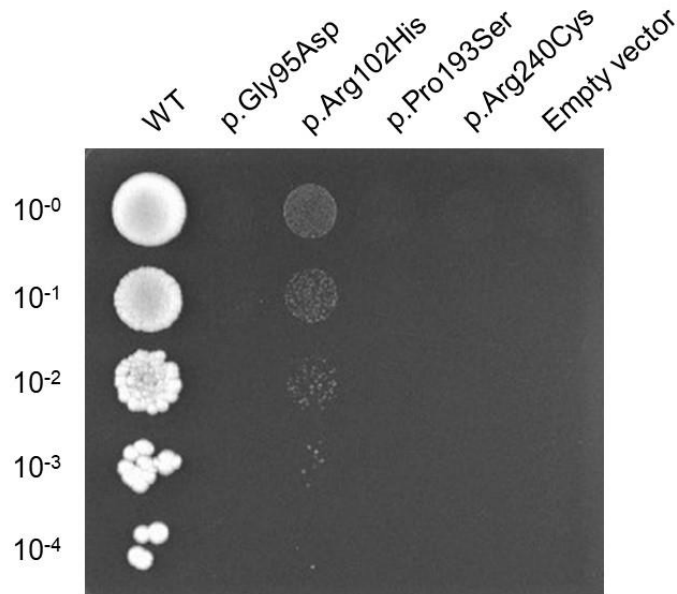

Figure 35. Yeast complementation assay for COQ4 mutations. Yeast lacking endogenous *COQ4* were transformed with a vector with no *COQ4* insert ('empty'), or with vectors containing wild-type or variant *COQ4.* Cultures were plated undiluted or diluted (1:10, 1:100, 1:1000, 1:10000) on media containing 5-FOA and grown at 30°C.

4.5.4 Studies in *Danio rerio*

Because of Coq$_{10}$ metabolism is critical to neuronal function (Spindler et al., 2009), we hypothesized involvement of *COQ4* in the neurodevelopment and used *Danio rerio* as a platform to test this possibility.

In situ hybridization studies showed that *coq4* mRNA is expressed in all the central nervous system from 24 hpf to 4 dpf stages of zebrafish development, and more specifically the expression is higher in the eye and brain and weaker in the trunk and tail at 5 dpf (Figure 37).

Neither zebrafish nor human *COQ4* cRNA overexpression impaired locomotion and morphology in embryos (Figure 36A, B and C). Over-expression in zebrafish embryos of cRNA bearing patients' variants (Figure D-N) (because of the under-expression of p.Arg102His mutation in patient's cDNA, it was not possible to test that mutation) indicated impairments: 1) in survival rate for

p.Pro193Ser allele (Figure 36G), 2) in coiling frequency for p.Gly95Asp and p.Pro193Ser alleles (Figure 36E and H), 3) in touch-evoked response for all (Figure 36F, I and N).
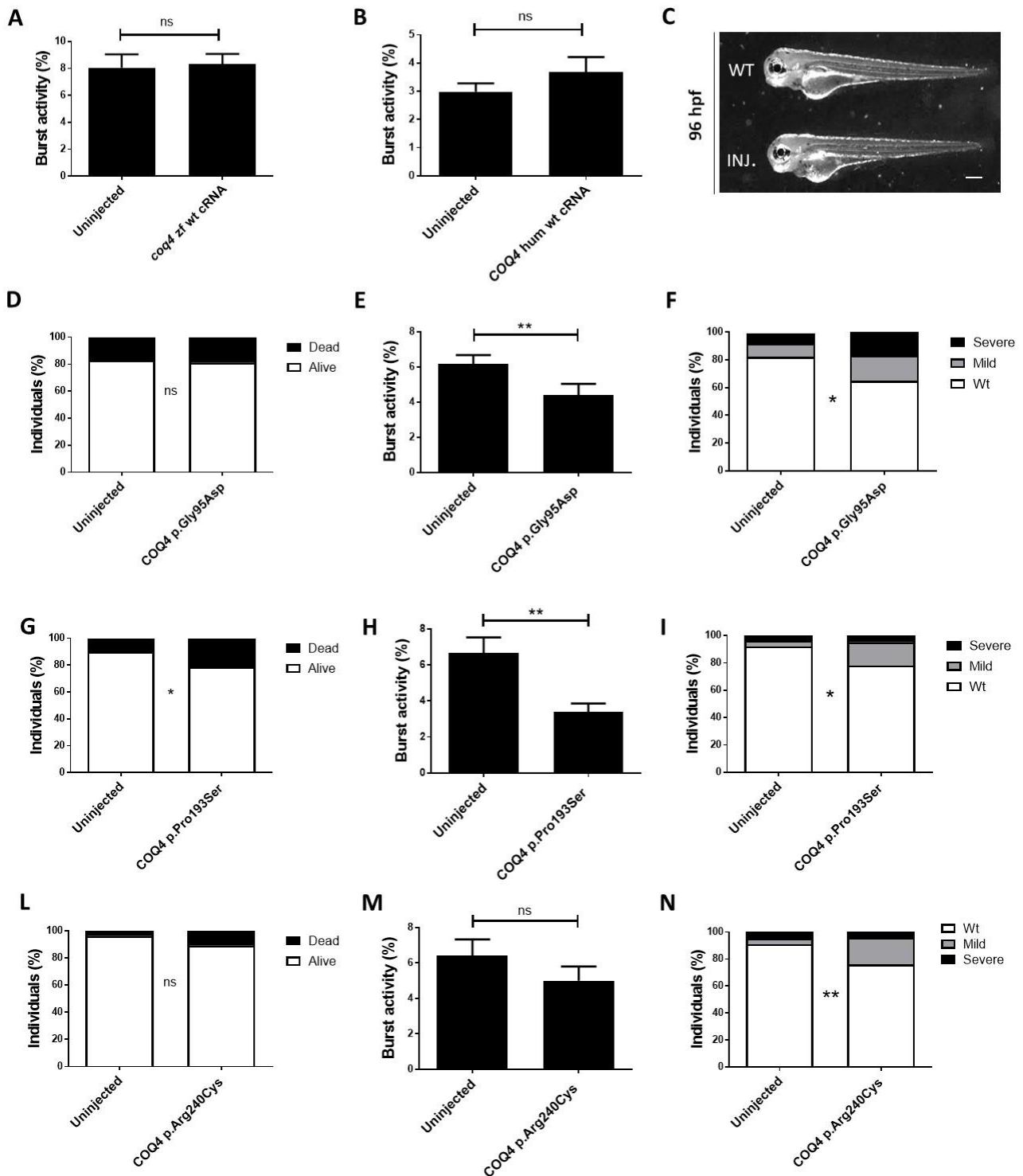
Figure 36. Overexpression in zebrafish. (A) *coq4* zebrafish wild type cRNA overexpression: coiling frequency. (B,C) *COQ4* human wild type cRNA overexpression: coiling frequency and morphology. Survival count, coiling frequency and touch-evoked response (in this order) are described for p.Gly95Asp (D, E, F), p.Pro193Ser (G, H, I) and p.Arg240Cys (L, M, N). Scale bar=300 µm.

Figure 37. coq4 mRNA localization in zebrafish embryo. Spatial-temporal expression pattern of *coq4* mRNA during zebrafish development. In situ hybridizations were performed using *coq4* RNA-antisense probe.

4.5.5 Studies in a *Danio rerio* knock-down model and generation of a knock-out model

Zebrafish *coq4* knock-down embryos were created using d-Cas9. Quantitative PCR revealed 90% *coq4* downregulation (Figure 38A), and significative differences were noticed when locomotion was tested by coiling frequency and touch-evoked response in 72 hpf embryos (Figure 38B and C).

Using CRISPR/Cas9 technology we generated *coq4* knock-out in NeuroD1 GcAMP6F background (Figure 39A). At the time of this writing, we have available data only in F0 KO embryos. In those, we showed significant differences in locomotion (distance and speed) (Figure 39B and 39C), similarly to d-Cas9 knock down model, with a slight rescue (though not statistically significant) using 1μM Idebenone, a synthetic analogue of CoQ_{10} (Figure 39B and 39C).

Whole-mount immunohistochemistry using anti-acetylated tubulin antibody, that binds stable microtubules, revealed an abnormal motor neuron growth at 5dpf (Figure 39D), whereas a partial reduction of the size of cerebellum not impacting amount of Purkinje cells was observed when anti-parvalbumin7 antibody, that labels these cells, was used (Figure 39E).
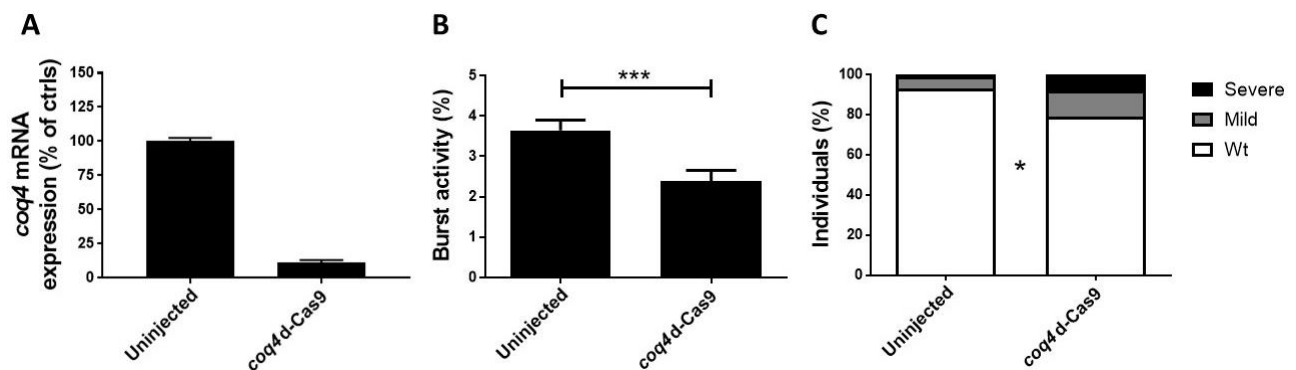
Figure 38. Legend in the next page.

70

Figure 38. coq4 d-Cas9 effects on zebrafish. (A) Measurement of *coq4* mRNA levels by quantitative real-time PCR. (B) Determination of coiling frequency and of (C) touch-evoked response.
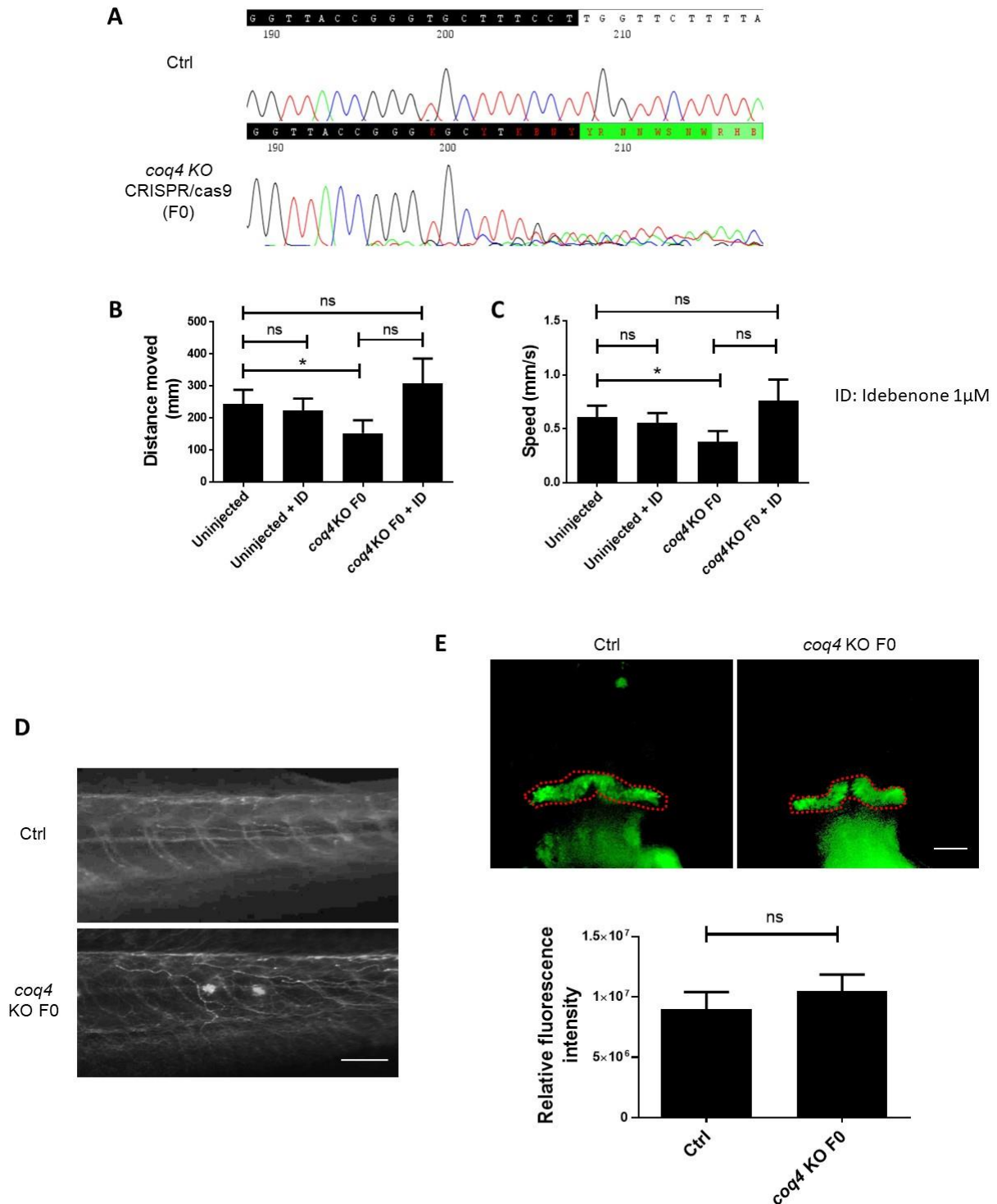
Figure 39. Generation and characterization of coq4 F0 KO zebrafish. (A) Genotyping of transformants zebrafish. (B,C) Locomotion assays with and without idebenone administration. (D) Whole-mount staining with anti-acetylated tubulin antibody labeling motor neurons (white filaments in the figure). (E) Staining with anti-parvalbumin7 antibody labeling cerebellum (red-dotted region). Scale bar in panel A=250 µm, scale bar in panel B=100 µm.

4.6 *STUB1* genetic studies

We found heterozygous mutation in *STUB1* in 7 index patients (Patient 54 p.Gly33Ser, patient 55 p.Tyr230Cysfs*9, patient 56 p.Pro228Ser, patient 57 p.Ala67Thr, patient 58 p.Arg225*, patient 59 p.Arg241Trp, patient 60 p.Pro57Leu). All mutations were rare or absent in gnomAD and our databases, predictably damaging *in silico* and segregating, whenever segregation studies were available, with the phenotype in familial cases (Figure 40). The clinical phenotype in affected individuals was consistent with adult onset complex syndrome characterized by both cerebellar ataxia and atrophy, cognitive-psychiatric disorders, and common presence of movement disorders and peculiar neuroimaging features. Besides these symptoms, also variable uncommon features (e.g. tremor, dystonia, urinary tract symptoms, hyperreflexia) occurred.

 Among the variants we found, all truncating mutations and two out of five missense laid in the ubiquitin ligase domain (U-box, Figure 41B), whereas three out of five missense were placed in the tetratricopeptide repeat (TRP) that binds HSP co-chaperones. All missense mutations are highly conserved (Figure 41B).
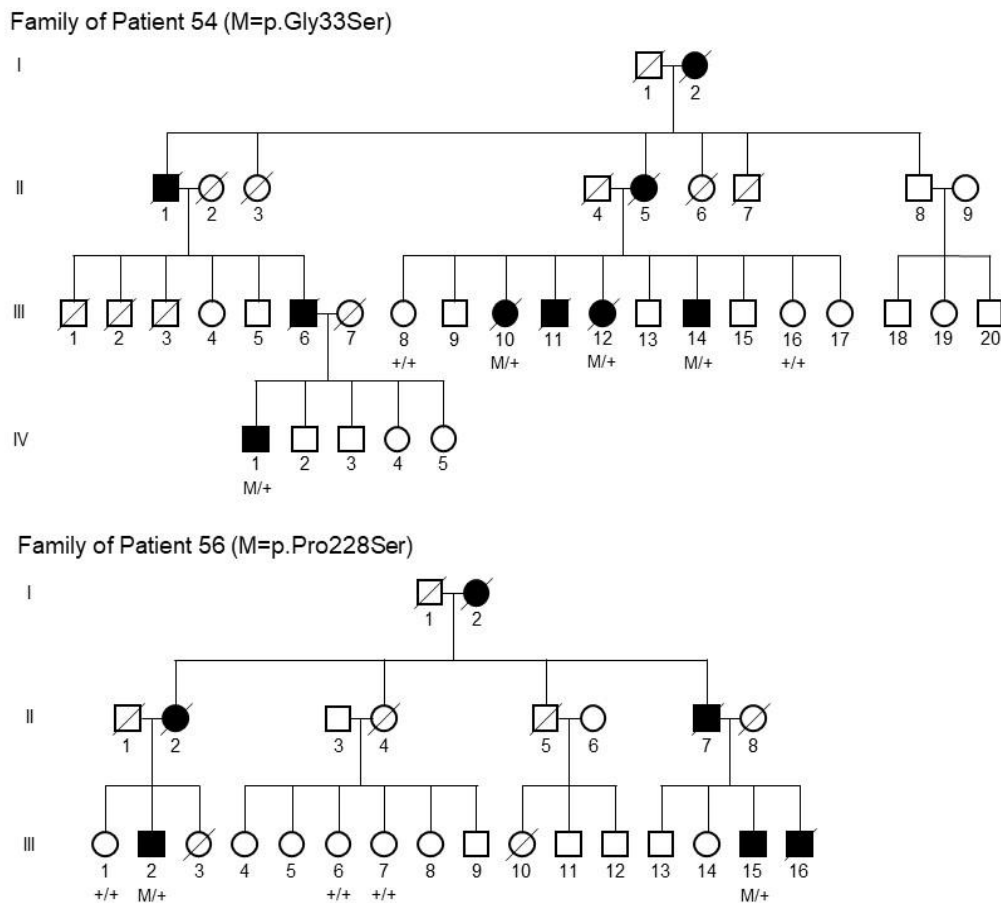
Figure 40. Continue in the next page.
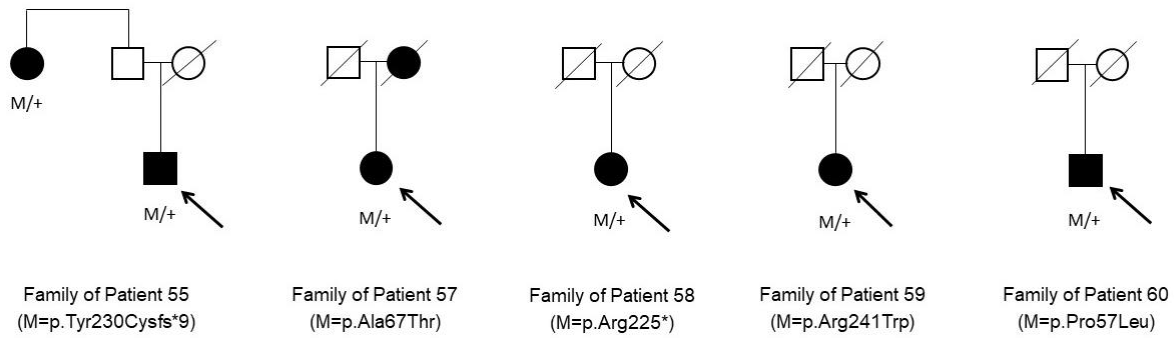
Figure 40. Pedigrees of families harbouring mutations in STUB1. Pedigrees of patient's 54 and 56 families indicate a clear autosomal dominant pattern of inheritance. Also pedigrees of patient's 55 and 57 families denote a familiar history of the disease, whereas other families result to be sporadic cases. "M" and "+" depict the presence of causative and wild-type alleles, respectively (pedigrees of the two large families are from De Michele et al. 2019, others are from Lieto et al. 2019).

Figure 41. Phylogenetic analysis of STUB1 missense mutations and scheme of all STUB1 pathogenetic variants so far described in SCA16 and SCA48. (A) Conservation analysis of missense variants found in our study. (B) All variants associated with autosomal recessive spinocerebellar ataxia 16 (SCAR16, not boxed) and autosomal dominant spinocerebellar ataxia 48 (SCA48, boxed) are depicted (UniProt Q9UNE7). Variants that we found in parallel with this NGS study (p.V264Gfs*4, p.P274Afs*3, p.L275Dfs*16), not listed among our positive cases, but published together (Lieto et al., 2019) were included in the figure. Novel SCA48 are indicated in bold and red boxes. p.L275Dfs*16 was cause of SCA48 in one of our patients (among the one not included among our positive cases) and already described by Genis and collaborators. The variant p.Y230Cfs*9 was also cause of SCA48 among our cohort, and it was previously described in SCAR16 in compound heterozygosity. Note that the c.612+1G>C and c*240T>C non-coding mutations (RefSeq NM_005861) described in SCAR16 kindred are not presented in this scheme. Legend: TPR=Tetratricopeptide repeat domains, U-box= Ubiquitin ligase domain. (part B is modified from Lieto et al., 2019).

4.7 *RNF216*, *COQ8A* and *ATP13A2* genetic outcomes

In four patients, NGS studies detected very rare variants in rare SCAR genes.

In patient 39 we found rare biallelic mutations in *RNF216* (namely, p.Met617Val; c.2061+3A>G) (Lieto et al., 2019). Segregation analyses showed a heterozygous status in the unaffected mother and brother (Figure 42A), whereas *in silico* predictions suggested a strong impact on protein function (p.Met617Val) and incorrect splicing mechanisms (c.2061+3A>G). Also phylogenetic analysis suggested that both methionine residue in position 617 and alanine in the splicing region are highly conserved among different species (Figure 42B). The phenotype was atypical showing a Huntington-like disease, severe cerebellar atrophy, late onset ataxia, normal peripheral nerve conduction velocities and absence of typical hypogonadism.
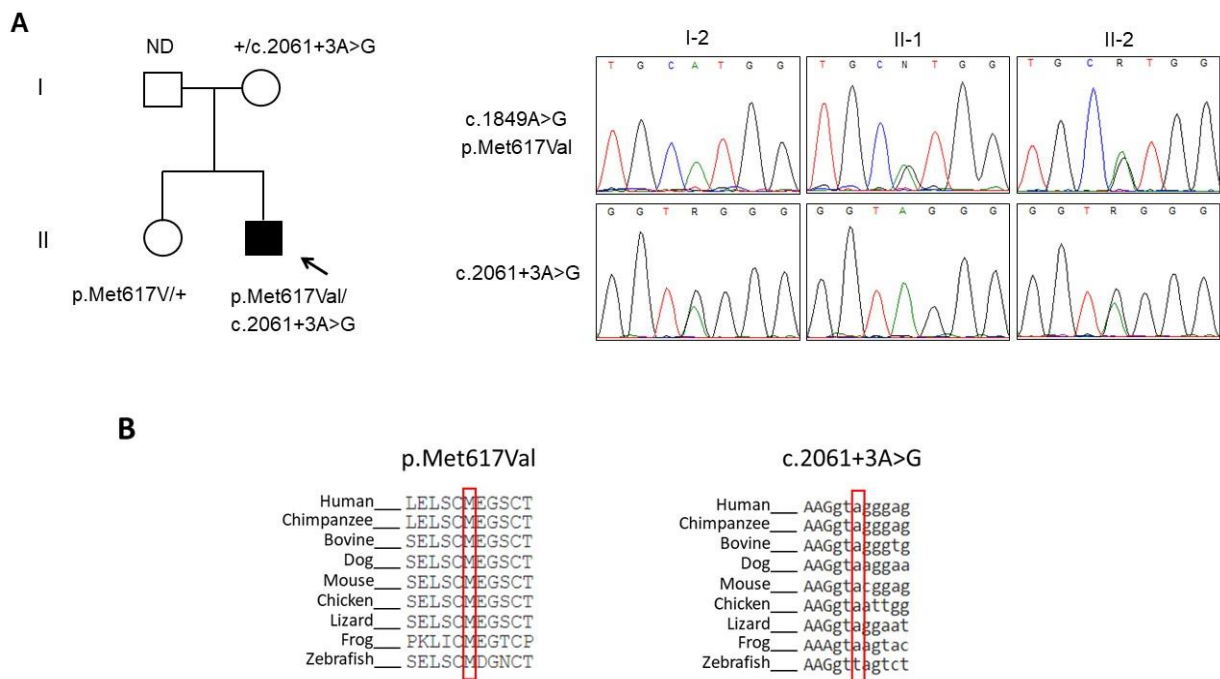
Figure 42. Patient's 39 family pedigree and RNF216 variants confirmations and phylogenetic analysis.

We found *COQ8A* pathogenic recessive variants in 2 patients (patients 12 and 13) presenting early handwriting impairment, dystonia and only mild ataxia. Patient 12 harbored a maternal missense (p.Gly615Asp) combined with a paternal c.589-3C>G variant whereas patient 13 carried a homozygous p.Arg348*. The p.Gly615Asp was predictably damaging *in silico*, highly phylogenetically conserved (Figure 43A) and described as pathogenic (Salviati et al, GeneReviews® 1993-2019). Furthemore, cDNA analysis demonstrated that only mRNA bearing p.Gly615Asp was expressed, thus suggesting that the mRNA carrying the c.589-3C>G variant was unstable or subjected to nonsense-mediated decay (Figure 43B).
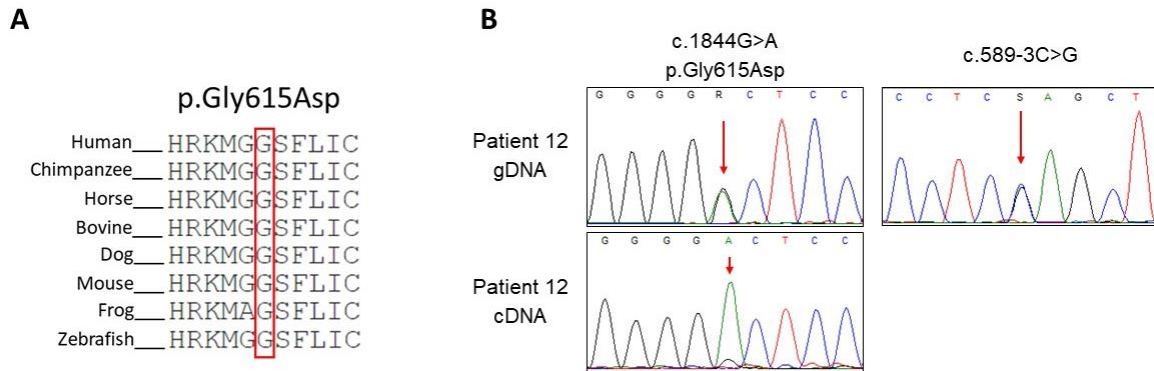
Figure 43. Phylogenetic analysis, COQ8A variants confirmations and cDNA mutation analysis of patient 12.

A novel homozygous mutation (p.Thr402Met, figure 44A) in *ATP13A2* was found in patient 7, a 65-year-old man who had developed gait ataxia at the age of 38 years and manifested "jerky movements" at age 55. The mutation was highly conserved (Figure 44B), located in an essential domain (E1-E2 ATPase domain) with a pivotal role for the enzymatic activity of the protein (Figure 44C), rare (MAF=2.85×10^{-5}) and potentially damaging but neither segregation studies nor functional investigations were possible.
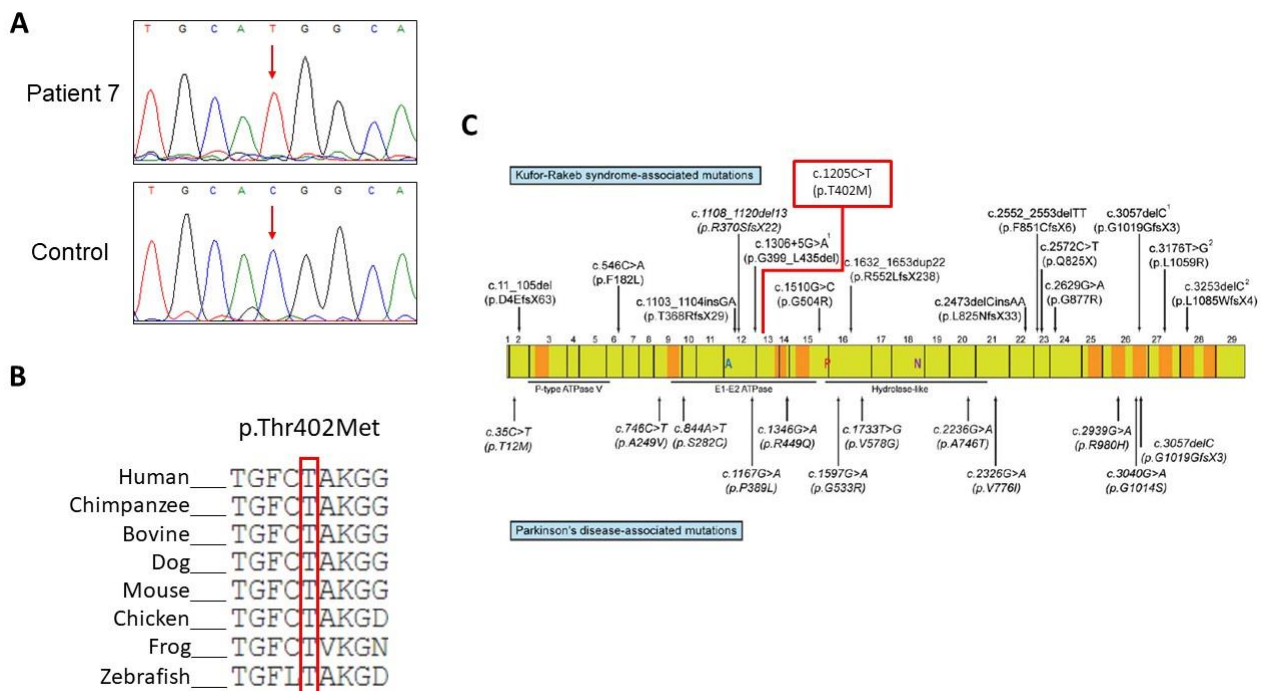
Figure 44. Novel mutation in ATP13A2. (A) Genetic confirmation. (B) Phylogenetic analysis. (C) Pathogenetic variants described in literature. The variant p.Thr402Met lay in a fundamental domain with ATPase activity, that is very important for protein function (panel A and B from Lieto et al. submitted, panel C is modified from Park et al., 2015).

5. Discussion

5.1 This NGS study broadened the genetic and phenotypic spectrum of hereditary ataxias

Among our patients, we found a diagnostic yield (25%) higher than the average described in literature (17%), but still lower that the one reported in application of exome sequencing as first-tier approach (36%) (Galatolo et al., 2018). Considering the very high coverage reached in multi-gene panels and their easier and faster analysis compared to exome sequencing, it is possible to argue that multi-gene panels are still worth to be used for quick screening of large cohorts. However, it must be said that at the time of the definition of this PhD project in 2016, we estimated from our experience that exome sequencing still had too high costs if compared to its low coverage not rarely insufficient to provide robust outcomes, and nevertheless subjected to wide gaps in coverage often filled by subsequent employment of Sanger sequencing. Thence, these considerations brought us to prefer multi-gene panel as first approach at that time, and to limit WES only in a second step.

To our knowledge, costs of both multi-gene panels and WES significatively reduced during last four years, and exome sequencing coverage quality improved. Hence, exome sequencing with about 100X average coverage is probably the most reasonable approach so far. Future reduction of technical costs for WGS should put the genome as first tier investigation.

In our cohort, *SPG7* represents the most common disease-causing gene, showing a similar high frequency as described in literature (Galatolo et al., 2018) (Coarelli et al., 2019) (Mancini et al., 2019). Also *PNPLA6,* a gene originally associated to HSP (Rainier et al., 2008) (Synofzik et al., 2014) (Wiethoff et al., 2016), was also common. Notably, pathogenic mutations in *STUB1*, a gene originally described in SCAR16 and recently associated to a dominant form (SCA48) (Genis et al., 2018), appear to have a relatively high frequency in our cohort. Finally, an elevate prevalence of mutations in genes coding for channels or their subunits confirms recent findings indicating their high occurrence among SCAs (Coutelier et al., 2017), whereas pathogenic variants in genes involved in metabolisms of complex lipids, cell structure, mitochondrial metabolism and DNA and RNA maintenance and repair seem even more common in SCARs, as widely previously stated (Synofzik and Schule, 2017).

However, despite a relatively high diagnostic rate, half of our patients resulted negative to NGS analysis, and 25% had an uncertain molecular diagnosis, percentages comparable to described exome sequencing analyses (Galatolo et al., 2018). In our case, the main conundrum was lack of detailed clinical information to refine genotype-phenotype associations, and limited access to family members due to the prevalently retrospective approach in our study. These data all suggest the

critical importance of full collection of blood samples in family members (at least the closest relatives) to perform accurate genetic analyses.

The presence of VOUS remains a challenge. In case of missense variants classified according to the ACMG criteria as score 3, only *in silico* tools cannot suffice and often family studies are partially conclusive. Therefore, it should be made effort to include functional analyses in the process of validating new variants to assess the pathogenic role of the putative mutations. Interestingly, SCAR genes usually not too rare were lower than expected (e.g. *SETX* and *ATM*) or even absent (e.g. *APTX* and *SACS*) in our cohort. This could happen since peculiar phenotypes strongly suggesting the involvement of a certain gene, as in AOAs, Ataxia-Telangectasia and ARSACS, usually undergo direct Sanger sequencing or the predominance of spastic paraplegia in those patients favors the use as first-tier of other HSP-related panels.

Two additional comments are worth mentioning. We identified rapidly a child bearing a *de novo* truncating mutation in *SLC2A1*, coding the major glucose transporter in brain (Baroni et al., 1992), and could be successfully treated with ketogenic diet. Although ultrarare, this is among the few HA cases that can be treated and reinforce the translational value of rapid gene identification. Second, we detected mutations associated with atypical phenotypes in rare SCAR genes (i.e. *AFG3L2*, *HARS*, *COQ4, COQ8A, RNF216*, *ATP13A2*) ending the diagnostic odyssey for these patients.

5.2 *SPG7* is the most common cause of hereditary ataxias in our cohort, and p.Ala510Val is a recurrent variant

Mutations in *SPG7* were the most common cause of HA in our cohort, confirming data in the literature and from others (van de Warrenburg et al., 2016) (Galatolo et al., 2018) (Coarelli et al., 2019). *SPG7* encodes for paraplegin, a ubiquitary expressed component of the mitochondrial AAA protease (m-AAA) in the inner mitochondrial membrane (Atorino et al., 2003) (Koppen et al., 2007). Paraplegin is highly expressed in Purkinje neurons (Sacco et al., 2010), pyramidal cortical neurons and spinal motor neurons (Patron et al., 2018). Paraplegin forms hetero-oligomeric complexes with AFG3L2, also involved in a dominant form of HA (SCA28, MIM# 610246) (Di Bella et al., 2010) and a recessive form of spastic ataxia (SPAX5, MIM# 614487) (Pierson et al., 2011). Paraplegin-AFG3L2 complex is involved in several pathways crucial for mitochondrial function, including mitochondrial protein quality control and homeostasis, and if compromised can lead to dysfunctions in mitochondrial protein synthesis, respiration, mitochondrial integrity, axonal transport and mitochondrial Ca^{2+} flux (Patron et al., 2018).

Notably, some of our patients harbored the common p.Ala510Val mutation either in homozygous or compound heterozygosity. Pathogenic role of this variants has been widely debated due to its high frequency, but studies in yeast confirmed its disease-causing role (Bonn et al., 2010). Moreover, recent large-cohort studies all agreed that p.Ala510Val is the most common mutation in SPG7 patients (Choquet et al., 2016) (Hewamadduma et al., 2018) (Mancini et al., 2018) (Coarelli et al., 2019) and influences clinical severity. Interestingly, statistical analyses in a large cohort of patients suggested that patients bearing the homozygous p.Ala510Val have a later onset and milder phenotype compared to patients bearing p.Ala510Val in compound heterozygous state (Mancini et al., 2018) (Coarelli et al., 2019). Furthermore, missense mutations (including p.Ala510Val) seems to be more commonly involved in cerebellar ataxia phenotype rather than spasticity, that seems to be often driven by complete loss-of-function mutations (Coarelli et al., 2019).

Normal parameters measured by OCR and normal levels of paraplegin expression in the single patient harboring the homozygous p.Ala510Val that we could investigate at the cellular level seems to support the presumed milder effect of this mutation. Nonetheless, the mutation seem to reduce AFG3L2 protein expression levels and this could be explained by the tight interaction between SPG7 and AFG3L2. How the specific variant impact on the heteromeric complex, structure relations or their interactions with other m-AAA mitochondrial proteases such as YME1L (Cesnekova et al., 2018) or even other mitochondrial proteins like OPA1 (Magri et al., 2018) involved in important mitochondrial dynamic process remain a future goal.

5.3 Drosophila *SPG7* plays a crucial role in neurons

Because of high frequency of *SPG7* as a disease-causing gene and the common occurrence of p.Ala510Val in paraplegin, we used *Drosophila melanogaster* to model *in vivo* paraplegin function and to set a new knock-in strain. When fly paraplegin downregulation occurred only in neurons, we observed a more aggressive phenotype than downregulation in other tissues. Silencing *dSPG7* in the whole fly was associated with ATP abnormal production and impaired vitality and locomotion similar to what we found in our KO model, and also comparable with mitochondrial defects of the already reported KO (Pareek et al., 2018) substantiating this model for future functional investigation when disease-related mutations appear questionable. Ongoing generation of a humanized p.Ala510Val knock-in model will clarify *in vivo* the role of this specific mutation and perhaps help to identify new approaches to its mechanisms in mice.

5.4 Bi-allelic mutations in *AFG3L2* cause an atypical form of autosomal recessive spastic ataxia type 5

In our study we described two patients with mutations in *AFG3L2*, one SCA28 patient and a SPAX5 child presenting an unusual milder phenotype characterized by stress-related episodic ataxia and loss of consciousness lasting few minutes without sequelae. AFG3L2 is a mitochondrial metalloprotease (m-AAA) and forms either hetero-oligomeric complexes with SPG7, or homo-oligomeric complex, conversely to SPG7 that exists only coupled with AFG3L2. SPG7-AFG3L2 complex is involved is several pathways crucial for mitochondrial function, as described before. AFG3L2 is selectively highly expressed in Purkinje cells (Patron et al., 2018), whereas SPG7 is highly expressed in Purkinje neurons as well (Sacco et al., 2010), but also in pyramidal cortical neurons and spinal motor neurons (Patron et al., 2018), a feature that could explain why mutations in *AFG3L2* lead to a more ataxic phenotype rather than *SPG7* in which loss-of-function mutations are more often related to spasticity. Moreover, studies in *AFG3L2* mice KO models showed structural abnormalities in mitochondria and neuronal loss (Almajan et al., 2012), and complete absence of AFG3L2 was seen to be lethal for neurons (Svenstrup et al., 2017). Despite molecular pathways underlying the involvement of AFG3L2 in conditions related to different patterns of inheritance are yet to be clarified, families harboring mutations in other genes (e.g. *KIF1A*, *SPG7*, *STUB1*) (Lee et al., 2015) (Sánchez-Ferrero et al., 2013) (De Michele et al., 2019) showed the similar issue of mono- and biallelic inheritance. Furthermore, also different disease-causing mechanisms seem to occur, since both dominant negative effect and haploinsufficiency have been described. (Di Bella et al, 2010). Nevertheless, as described for *SPG7*, tight interaction between SPG7 and AFG3L2 make functional studies even more challenging for several abovementioned reasons.

In our patient we detected a frameshift mutation (p.Val212Glyfs*4) subjected to nonsense-mediated decay and a missense mutation (p.Val723Met) predicted *in silico* to have a pathogenetic significance. Remarkably, p.Val723Met was previously described (Tunc et al., 2019) in compound heterozygosity with p.Tyr616Cys, a mutation already outlined as cause of SPAX5 when in homozygosity (Pierson et al., 2011) and whose disease-causing role has been recently confirmed (Tulli et al., 2019), hence corroborating the hypothesis that p.Val723Met has a pathogenic effect, and that together with the clearly pathogenic p.Val212Glyfs*4, are the disease-causing mutations in our patient.

5.5 Drosophila *AFG3L2* has a pivotal role in development

dAFG3L2 silencing in the whole fly led to a severe phenotype with shortened lifespan and locomotor and biochemical abnormalities, whereas specific downregulation in muscles brought to a more severe phenotype corroborated by histological abnormalities in mitochondria and fibres structures. An even nearly-lethal phenotype was observed when *dAFG3L2* was silenced in neurons. Bypassing the *dAFG3L2* silencing in all developmental stages with the use of the GAL80^{ts} application of GAL4-UAS system, we could downregulate *dAFG3L2* in neurons of adult flies and observe a healthier phenotype. These data suggest that AFG3L2 is critical for neurodevelopment (at least in flies), a finding in apparent contrast with the relatively less aggressive effects seen when other mitochondrial AAA-proteases were knocked-out in Drosophila (paraplegin in this study and in Pareek et al. 2018; YME1L in Qi et al. 2016). All these KO flies present premature aging, progressive locomotor deficiency, neurodegeneration, mitochondrial defects and increased susceptibility to oxidative stress, in a similar but less severe manner compared to what we observed in the *dAFG3L2* knock-down model, so that we could speculate on the possibility that in the m-AAA protein complex the essential subunit is AFG3L2.

5.6 Bi-allelic mutations in histidyl-tRNA synthetase (*HARS*) cause a novel recessive ataxia-related phenotype

We identified three patients in two unrelated families carrying bi-allelic *HARS* variants present with novel clinical features. Histidyl-tRNA synthetase (*HARS*) is a homodimeric class IIa aminoacyl tRNA synthetase that charges tRNA with the amino acid histidine in the cytoplasm. Aminoacyl-tRNA synthetases are ubiquitously expressed and highly conserved enzymes that catalyze the conjugation of tRNA to cognate amino acids in the cytoplasm, in mitochondria, or in both locations (Antonellis and Green, 2008). Mutations in *HARS* were found to cause two different disorders through either autosomal dominant or autosomal recessive inheritance, a feature shared with only four other ARS enzymes (Meyer-Schuman and Antonellis, 2017). Puffenberger and collaborators (Puffenberger et al, 2012) identified a homozygous missense mutation in three Old Order Amish patients presenting Usher Syndrome type IIIB syndrome (MIM# 614504), an autosomal recessive disorder characterized by motor impairment, and progressive vision and hearing loss in childhood. Later, several studies (Vester et al., 2013) (Safka Brozkova et al., 2015) (Abbot et al., 2018) (Royer-Bertrand et al., 2019) revealed monoallelic mutations in *HARS* as the cause of dominantly inherited Charcot-Marie-Tooth disease type 2W (MIM# 616625), a motor and sensory peripheral neuropathy affecting both lower and upper limbs.

Microcephaly, mild-to severe intellectual disability, skeletal deformities, ataxic gait and clinical features affecting the cerebellar and pyramidal tract systems are common to the three patients. Choreo-athetoid movements with ballismus in patient 16 and dystonic postures in the two sisters characterize this novel ataxia-related recessive phenotype. Hearing loss was found in one case and peripheral nerves involvement —the hallmark of CMT2W— was minor feature and occurred only in the two sisters. Retinal dysfunction, typical of *HARS*-related Usher Syndrome IIIB, was not found in patients. Brain MRI was similar in the extent of microcephaly and moderate enlargement of cerebellar folia but with peculiar features in patient 16.

Several evidences suggest that the mutations are pathogenetic. First, the rather similar clinical features seen in patients. Second, missense mutations were predicted deleterious by most *in silico* algorithms. Both qPCR and Western blotting revealed a significant reduction of *HARS* mRNA and protein expression, suggesting loss of enzyme function. Third, functional studies in yeast confirmed the loss-of-function of p.Val244Cysfs*6 and p.Leu305dup whereas we observed that missense mutations did not affect the growth of colonies. Studies reported by others widely demonstrated that ARS alleles incapable to complement yeast orthologues are also unable to charge tRNAs, reinforcing *S. cerevisiae* as an efficient model of human mutations in tRNA synthetases genes (Antonellis et al., 2006) (Jordanova et al., 2006) (Nangle et al., 2007) (McLaughlin et al., 2012) (Vester et al., 2013). The fact that all our patients harbor at least a mutation that is not coherent with a completely null function is not surprising, considering the crucial role of ARSs enzymes that requires a residual activity for viability (Meyer-Schuman and Antonellis, 2017). Fourth, we corroborated the hypothesis of a loss of function by studying aminoacylation assay in patients' primary cells where we found a strong and significant reduction of tRNA charging suggestive of a severe yet not complete reduction of the protein function.

Overall, our data expand the clinical manifestations of *HARS* and enlarge the set of aminoacyl-tRNA synthetases associated with ataxia and movement disorders. Most ARSs are involved in a large number of genetic diseases causing impairments in different tissues (Meyer-Schuman and Antonellis, 2017) (Royer-Bertrand et al., 2019). Both loss- and gain-of-function mechanisms have been proposed (Wallen and Antonellis, 2013), but impairment in central nervous system is likely due to aminoacylation decrease in tissues with high protein synthesis demands such as neurons, not involving other tissues that may require only a basal or housekeeping level of protein synthesis (Royer Bertrand et al., 2019). Further canonical and non-canonical functions of ARSs could also be involved (Fuchs et al., 2019).

5.7 *COQ4* mutations cause early ataxia and developmental delay

In our cohort, we identified two patients harboring biallelic variants in *COQ4*. *COQ4* encodes a protein involved in CoQ_{10} biosynthesis, whose primary deficiency encompass a heterogeneous group of mitochondrial disorders caused by mutations in ten genes (*COQ2*, *COQ4*, *COQ5*, *COQ6*, *COQ7*, *COQ8A*, *COQ8B*, *COQ9*, *PDSS1*, *PDSS2*) resulting in five major phenotypes: encephalomyopathy, severe infantile multi-systemic disease, cerebellar ataxia, isolated myopathy, and steroid-resistant nephrotic syndrome (Emmanuele et al., 2012) (Laredj et al., 2014) (Desbats et al., 2015) (Freyer et al., 2015) (Wang et al., 2017) (Malicdan et al., 2018). Moreover, different phenotypes are associated with mutations in the same CoQ genes, suggesting that CoQ_{10} residual level may influence differences of phenotypes, as described in yeast (Desbats et al., 2016).

The exact role of *COQ4* is still largely unknown and its contribution to CoQ_{10} biosynthesis is questioned, indeed studies in yeast suggest a structural role in the multi-protein complex involved in CoQ biosynthesis (Tran and Clarke, 2007) (Marbois et al., 2009). According to reports of patients harboring *COQ4* mutations (Salviati et al., 2012) (Brea-Calvo et al., 2015) (Chung et al., 2015) (Sondheimer et al., 2017) (Bosch et al., 2018) (Lu et al., 2019) (Caglayan et al., 2019), respiratory distress, cerebellar atrophy, lactic acidosis, and decreased levels of both CoQ_{10} and respiratory chain CII+CIII enzyme activity are the most common features in humans, with death occurring usually after birth or in early childhood. Recently, mutations in *COQ4* have been described in fatal infantile ataxia (Caglayan et al., 2019). Our patients had a longer survival and a relatively milder though chronically disabling phenotype, with no decreased CoQ_{10} levels and impaired mitochondrial respiratory chain activities. Studies in yeast and zebrafish suggested a deleterious effect for all missense variants, whereas investigations in *COQ4* knock-down and knock-out zebrafish indicated a locomotor phenotype and a potential role in not-canonical COQ4 function like neuronal development. However, considering that we performed our studies in F0 embryos and that Idebenone administration provided unclear results, further experiments in stable KO model will be necessary.

5.8 Additional variants in HA genes expand the spectrum of clinical phenotypes

5.8.1 *STUB1* heterozygous mutations are a common cause of SCA48

Detection of pathogenetic mutations in *STUB1* in SCA48 (MIM# 618093) patients (De Michele et al., 2019) (Lieto et al., 2019) further expanded our array of novel genotypes and phenotypes adding *STUB1* to the growing list of spastic-ataxia gene no longer respecting the zigosity boundaries.

STUB1 was originally associated with autosomal recessive spinocerebellar ataxia 16 (SCAR16, MIM# 615768), a neurodegenerative disorder characterized by early-onset cerebellar ataxia and cerebellar atrophy with cognitive impairment and childhood onset, variably associated to several clinical signs like spasticity, epilepsy, myoclonus, tremor, urinary tract symptoms and peripheral neuropathy, hence resulting in different clinical pictures like spastic ataxia, Gordon-Holmes syndrome and multisystemic neurodegeneration (Shi et al., 2013) (Cordoba et al., 2014) (Synofzik et al., 2014) (Depondt et al., 2014) (Heimdal et al., 2014) (Shi et al., 2014) (Bettencourt et al., 2015) (Kawarai et al., 2016) (Hayer et al., 2017) (Gazulla et al., 2018) (Turkgenc et al., 2018), and only three reports have described SCA48 kindred presenting adult onset cerebellar cognitive-affective syndrome (CCAS), sometimes preceding motor cerebellar signs, and often mimicking Huntington-like features (Genis et al., 2018) (De Michele et al., 2019) (Lieto et al., 2019).

STUB1 encodes CHIP (Carboxy terminus Hsp70 Interacting Protein), protein with roles as chaperone, co-chaperone and ubiquitin ligase (Ballinger et al., 1999) (Zhang et al., 2005). CHIP comprises an N-terminal tetratricopeptide repeat (TPR) domain required for binding to HSP70 and HSP90 chaperones, a coiled-coiled domain involved in dimerization, and a C-terminal U-box domain important for ubiquitin ligase function. CHIP is indeed believed to participate in protein quality control by targeting a broad range of chaperone protein substrates for degradation. Why mutations in *STUB1* lead to the complex phenotype observed in SCAR16 and SCA48 remains to be clarified but it is possible that the pathogenetic mechanism implies dysfunction of autophagy pathway. CHIP regulates autophagy flux via its function in targeting phosphorylated transcription factor EB (TFEB) for degradation or controlling regulation of mitochondrial biogenesis via reduction in the master gene PGC1alpha, or both (Rao et al., 2017) (Sha et al., 2017). Also, a *Stub1$^{-/-}$* mouse (Wei et al., 2015) shows findings suggestive of impaired autophagy. Our patients presented several features shared with SCAR16 but it remains puzzling why healthy parents in SCAR16 kindred are healthy in spite of having a single mutant allele. It seem unlikely that type and location of mutation may influence the final phenotype, since all so far described mutations lay in all domains, but is noteworthy that a large number of variants, especially truncating mutations, are located in the U-box domain, maybe suggesting a possible association between the position of disease-causing variants and the phenotype developed or even only certain clinical signs. This approach was tested in a recent meta-analysis (Madrigal et al., 2019) indicating that point mutations in U-box domain have severe effects on the overall CHIP functions and are tightly related to cognitive dysfunction in SCAR16 patients, hence proposing that mutations in certain CHIP domains could bring to specific phenotypic manifestations. The limited number of monoallelic missense

variants and the lack of functional investigations on protein function and structural features prevented similar speculations of a genotype-phenotype correlations in SCA48.

The role of additional factors, including epigenetics and modifier factors, should be investigated and might provide explanation about the reason why parents of SCAR16 patients are not affected. Using eleven bioinformatic tools (Table S3 and Table S4 in supplemental material), we corroborated *in silico* the hypothesis that the pathogenicity of mutations causing SCA48 in our patients is comparable to disease-causing mutations so far described in *STUB1*. Because of the lack of biological samples beyond peripheral blood for all these patients, we could not set up any functional work-up or study in animal models for our SCA48 variants, a pipeline that we could instead perform whenever other biological samples (e.g. patient's derived skin fibroblasts) were available to perform studies preliminary to investigations in animal models as we did for other genes (i.e. *SPG7, AFG3L2, HARS, COQ4*). Thence, functional studies in SCA48 are much needed and our *in silico* studies were preliminary.

5.8.2 Two novel mutations in *RFN216* are associated with ataxia, chorea and fertility but without hypogonadism

We found pathogenetic biallelic mutations in *RNF216* in a patient with an atypical *RNF216*-related phenotype in the absence of hypogonadism (Lieto et al., 2019). *RNF216* encodes an E3 ubiquitin ligase that attaches ubiquitin to protein substrate marking them for a proteasome-mediated degradation, and contains RING (Really Interesting New Gene) finger domain 1, IBR (In Between Ring domain) and RING finger domain 2 (Calandra et al., 2019). The mutations so far described are spread throughout the whole gene, thus suggesting that no particular hotspots regions occur. *RNF216* is structurally similar to parkin, an E3 ubiquitin ligase that is muted in a recessive form of Parkinson disease. *RNF216* mutations alone or combined with *OTUD4* mutations have initially been associated to ataxia with cerebellar atrophy and hypogonadotropic hypogonadism (Gordon-Holmes syndrome) and dementia (Margolin et al., 2013). The phenotype successively widened to include chorea with behavioral problems (Santens et al., 2015), and possibly features of the 4H syndrome (Hypomyelination, Hypodontia, Hypogonadotropic Hypogonadism) (Ganos et al., 2015). MRI usually shows supratentorial diffuse white matter changes. Seventeen patients from ten families have been so far reported (Margolin et al., 2013) (Sawyer et al., 2014) (Santens et al., 2015) (Ganos et al., 2015) (Alqwaifly and Bohlega, 2016) (Calandra et al., 2019). Our findings contribute to the clinical and genetic heterogeneity of *RNF216*-related syndromes.

5.8.3 Recessive mutations in *COQ8A* are associated with dystonia-ataxia with early handwriting deterioration

We found disease-causing mutation in *COQ8A* in two patients of our cohort, a gene with an essential role as an electron carrier and endogenous antioxidant. The atypical presentation of our patients with prominent writing deterioration, possibly representing the initial manifestation of cerebellar disease, contributed to expand the clinical and biochemical manifestations of the CoQ_{10} deficiency syndromes (Galosi et al., 2019).

5.8.4 A homozygous mutation in *ATP13A2* is associated with progressive myoclonic ataxia

We detected a homozygous mutation in *ATP13A2* associated with an uncommon phenotype related to this gene. Our patient with progressive myoclonus ataxia, parkinsonism and both cerebellar and cortical atrophy resembles at large patients with Kufor-Rakeb Syndrome (KRS, MIM# 606693) caused by mutations in *ATP13A2* (Najim al-Din et al., 1994) (Ramirez et al., 2006). To date, more than 30 KRS individuals in 11 families have been described (Park et al., 2015) and our additional case broadens the phenotype since action myoclonus has been reported once (Rohani et al., 2017) and ataxia is rare (Eiberg et al, 2012) in KRS.

ATP13A2 belongs to the P-type superfamily of ATPases that transport inorganic cations and other substrates across cell membranes (Schultheis et al., 2004), and is involved in Mn^{2+} and more importantly in Zn^{2+} metabolism (Gitler et al., 2009) (Kong et al., 2014), mitochondrial bioenergetics (Gusdon et al., 2012) (Ramonet et al., 2012) (Grunewald et al., 2012), and the autophagy–lysosomal pathway (Dehay et al., 2012) (Usenovic et al., 2012) (Park et al., 2015). Noteworthy, mutation we found in our patient lays in a domain crucial for the enzymatic catalytic activity (Park et al., 2015). Morerover, heterozygous mutations in *ATP13A2* have also been identified in early-onset Parkinson's disease (PD), suggesting that these mutations could be a risk factor or an age-of-onset modifier in Parkinson's disease (Di Fonzo et al., 2007) (Lin et al., 2008). Interestingly, recessive mutations in *ATP13A2* have been reported as cause also of neuronal ceroid lipofuscinoses (NCLs) (Bras et al., 2012).

6. Conclusions and future perspectives

Despite more than forty years has passed since Frederick Sanger revolutionized the entire landscape of biology with the development of a "smart" system for DNA sequencing, his method is still a milestone for several applications. Nevertheless, modern clinical genetics face the huge level of heterogeneity and requires more massive, high-throughput approaches to solve the unsolved cases. An additional advantage is the demolition of boundaries such as the zygosity (dominant and recessive variants can occur in the same gene) or the clinical classification (no longer applicable when several phenotypes appear part of the same spectrum of disorders) and also the possibility to identify additional minor variants acting as potential modifiers (a research field not fully explored thus far). Moreover, different applications of NGS could also lead to new approaches, such as transcriptomics and epigenetics, driving to the discovery of the involvement of few similar pathological mechanisms in many different genetic forms. These issues could change the future approach in the research of drugs and treatments, shifting focus from single disorders to hubs of molecular pathways, hence moving the attention more on the biological processes at the base of diseases rather than the symptoms.

Which NGS application suits better clinical needs is an argument still debated. Despite methodological and technical advantages of multi-gene panels, exome sequencing gives the undeniable advantage of searching in the full coding regions, and offers the possibility to create *in silico* gene panels customizable as research information brings to light new players in HAs. Surely, as costs are running down and bioinformatics tools become more robust and easy-to-use, additional platforms such as the full genome (once in a lifetime test) should be consider provided that data are continuously revised and interpreted in combination with functional studies and natural history information.

However, *in vivo* models remains essential tools to validate the role of VOUS and to explore the role of genes in biological pathways and their involvement in pathological processes.

To conclude, our study underlies the impact of NGS in HAs, but also the significant role as a crucial hub between clinical and research studies. Definition of NGS technologies in an integrated approach with detailed clinical and functional investigations can take the best from the massive technology and also permit to discover new genes and novel pathways in HAs. This is fundamental to generate faster translation to personalized therapies.

7. Contributions and acknowledgements

8. References

Abbott, J.A. *et al.* Substrate interaction defects in histidyl-tRNA synthetase linked to dominant axonal peripheral neuropathy. *Hum. Mutat.* 39, 415–432 (2018).

Abele, M. *et al.* The aetiology of sporadic adult-onset ataxia. *Brain* 125, 961–968 (2002).

Almajan, E. *et al.* AFG3L2 supports mitochondrial protein synthesis and Purkinje cell survival. *J. Clin. Invest.* 122, 4048–4058 (2012).

Alqwaifly, M. & Bohlega, S. Ataxia and hypogonadotropic hypogonadism with intrafamilial variability caused by *RNF216* mutation. *Neurol. Int.* 8, 6444–6446 (2016).

America, N. *et al.* Truncating mutations in the last exon of *NOTCH2* cause a rare skeletal disorder with osteoporosis. *Nat. Genet.* 43, 306–308 (2011).

Anheim, M., Tranchant, C. & Koening, M. The Autosomal Recessive Cerebellar Ataxias. *N. Engl. J. Med.* 366, 636–646 (2012).

Antonellis, A. *et al.* Functional Analyses of Glycyl-tRNA Synthetase Mutations Suggest a Key Role for tRNA-Charging Enzymes in Peripheral Axons. *J. Neurosci.* 26, 10397–10406 (2006).

Antonellis, A. & Green, E.D. The Role of Aminoacyl-tRNA Synthetases in Genetic Diseases. *Annu. Rev. Genomics Hum. Genet.* 9, 87–107 (2008).

Applications of Clinical Microbial Next-Generation Sequencing: Report on an American Academy of Microbiology Colloquium held in Washington, DC, in April 2015

Atorino, L. *et al.* Loss of m-AAA protease in mitochondria causes complex I deficiency and increased sensitivity to oxidative stress in hereditary spastic paraplegia. *J. Cell Biol.* 163, 777–787 (2003).

Auburger, G. *et al.* Primary Skin Fibroblasts as a Model of Parkinson's Disease. *Mol. Neurobiol.* 46, 20–27 (2012).

Bae, Y. *et al.* Anatomy of zebrafish cerebellum and screen for mutations affecting its development. *Dev. Biol.* 330, 406–426 (2009).

Ballinger, C.A. *et al.* Identification of CHIP, a Novel Tetratricopeptide Repeat-Containing Protein That Interacts with Heat Shock Proteins and Negatively Regulates Chaperone Functions. *Mol. Cell. Biol.* 19, 4535–4545 (1999).

Baroni, M.G. *et al.* Polymorphisms at the *GLUT1* (HepG2) and *GLUT4* (muscle/adipocyte) glucose transporter genes and non-insulin-dependent diabetes mellitus (NIDDM). *Hum. Genet.* 88, 557–561 (1992).

Barrangou, R. *et al.* CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* 315, 1709–1712 (2007).

Barresi, S. *et al.* Mutations in the IRBIT domain of *ITPR1* are a frequent cause of autosomal dominant nonprogressive congenital ataxia. *Clin. Genet.* 91, 86–91 (2017).

Bettencourt, C. *et al.* Clinical and Neuropathological Features of Spastic Ataxia in a Spanish Family with Novel Compound Heterozygous Mutations in *STUB1*. *Cerebellum* 14, 378–81 (2015).

Blaser, R.E. & Vira, D.G. Experiments on learning in zebrafish (*Danio rerio*): A promising model of neurocognitive function. *Neurosci. Biobehav. Rev.* 42, 224–231 (2014).

Bohles, H. *et al.* Hyperinsulinaemic hypoglycaemia--leading symptom in a patient with congenital disorder of glycosylation Ia (phosphomannomutase deficiency). *J. Inherit. Metab. Dis.* 24, 858–862 (2001).

Bonn, F., Pantakani, K., Shoukier, M., Langer, T. & Mannan, A.U. Functional Evaluation of Paraplegin Mutations by a Yeast Complementation Assay. *Hum. Mutat.* 31, 617–621 (2010).

Bosch, A.M. *et al.* Coenzyme Q_{10} deficiency due to a *COQ4* gene defect causes childhood-onset spinocerebellar ataxia and stroke-like episodes. *Mol. Genet. Metab. Reports* 17, 19–21 (2018).

Bosco, G., Campbell, P., Leiva-Neto, J.T. & Markow, T.A. Analysis of Drosophila Species Genome Size and Satellite DNA Content Reveals Significant Differences Among Strains as Well as Between Species. *Genetics* 177, 1277–90 (2007).

Brand, A.H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* 118, 401–415 (1993).

Bras, J., Verloes, A., Schneider, S.A., Mole, S.E. & Guerreiro, R.J. Mutation of the parkinsonism gene *ATP13A2* causes neuronal ceroid-lipofuscinosis. *Hum. Mol. Genet.* 21, 2646–2650 (2012).

Brea-Calvo, G. *et al. COQ4* Mutations Cause a Broad Spectrum of Mitochondrial Disorders Associated with CoQ_{10} Deficiency. *Am. J. Hum. Genet.* 96, 309–317 (2015).

Caglayan, A. O. *et al. COQ4* Mutation Leads to Childhood-Onset Ataxia Improved by CoQ_{10} Administration. *Cerebellum* 18, 665–669 (2019).

Cagnoli, C. *et al.* Large Pathogenic Expansions in the SCA2 and SCA7 Genes Can Be Detected by Fluorescent Repeat-Primed Polymerase Chain Reaction Assay. *J. Mol. Diagnostics* 8, 128–132 (2006).

Calandra, C.R. *et al.* Gordon Holmes Syndrome Caused by *RNF216* Novel Mutation in 2 Argentinean Siblings. *Mov. Disord. Clin. Pract.* 6, 259–262 (2019).

Campuzano, V. *et al.* Friedreich's Ataxia: Autosomal Recessive disease Caused by an Intronic GAA Triplet Repeat Expansion. *Science* 271, 1423–1427 (1996).

Casari, G. *et al.* Spastic Paraplegia and OXPHOS Impairment Caused by Mutations in Paraplegin, a Nuclear-Encoded Mitochondrial Metalloprotease. *Cell* 93, 973–983 (1998).

Caygill, E.E. & Brand, A.H. The GAL4 System: A Versatile System for the Manipulation and Analysis of Gene Expression. *Methods Mol. Biol.* 1478, 33–52 (2016).

Cesnekova, J., Rodinova, M., Hansikova, H., Zeman, J. & Stiburek, L. Loss of Mitochondrial AAA Proteases AFG3L2 and YME1L Impairs Mitochondrial Structure and Respiratory Chain Biogenesis. *Int. J. Mol. Sci.* 19, 1–13 (2018).

Choquet, K. *et al. SPG7* mutations explain a significant proportion of French-Canadian spastic ataxia cases. *Eur. J. Hum. Genet.* 24, 1016–1021 (2016).

Chung, W. K. *et al.* Mutations in *COQ4*, an essential component of coenzyme Q biosynthesis, cause lethal neonatal mitochondrial encephalomyopathy. *J. Med. Genet.* 52, 627–635 (2015).

Coarelli, G. *et al.* Loss of paraplegin drives spasticity rather than ataxia in a cohort of 241 patients with SPG7. *Neurology* 92, e2679–e2690 (2019).

Coarelli, G., Brice, A. & Durr, A. Recent advances in understanding dominant spinocerebellar ataxias from clinical and genetic points of view. *F1000 Res.* 7, 1–10 (2018).

Cordoba, M., Rodriguez-Quiroga, S., Gatto, E., Alurralde, A. & Kauffman, M. Ataxia plus myoclonus in a 23-year-old patient due to *STUB1* mutations. *Neurology* 83, 287–288 (2014).

Cortese, A. et al. Biallelic expansion of an intronic repeat in *RFC1* is a common cause of late-onset ataxia. *Nat Genet*. 51(4):649-658 (2019).

Coutelier, M. *et al.* A Recurrent Mutation in *CACNA1G* Alters and Causes Autosomal-Dominant Cerebellar Ataxia. *Am. J. Hum. Genet.* 97, 726–737 (2015).

Coutelier, M. *et al.* A panel study on patients with dominant cerebellar ataxia highlights the frequency of channelopathies. *Brain* 140, 1579–1594 (2017).

Curran, B.P.G. & Bugeja, V. Basic Investigations in *Saccharomyces cerevisiae*. *Methods Mol. Biol.* 1163, 1–14 (2014).

D'Amore, A. *et al*. Next Generation Molecular Diagnosis of Hereditary Spastic Paraplegias: An Italian Cross-Sectional Study. Front. Neurol. 9, 981 (2018).

D'Costa, A. & Shepherd, I.T. Zebrafish Development and Genetics: Introducing in a Large Introductory Laboratory Class. *Zebrafish* 6, 169–177 (2009).

De Michele, G. *et al*. Spinocerebellar ataxia 48 presenting with ataxia associated with cognitive, psychiatric, and extrapyramidal features: A report of two Italian families. *Parkinsonism Relat. Disord.* 65, 91-96 (2019).

Dehay, B. *et al.* Loss of P-type ATPase ATP13A2/PARK9 function induces general lysosomal deficiency and leads to Parkinson disease neurodegeneration. *Proc. Natl. Acad. Sci.* 109, 9611–9616 (2012).

Depondt, C. *et al.* Autosomal recessive cerebellar ataxia of adult onset due to *STUB1* mutations. *Neurology* 82, 1749–1750 (2014).

Desbats, M. A., Lunardi, G., Doimo, M., Trevisson, E. & Salviati, L. Genetic bases and clinical manifestations of coenzyme Q10 (CoQ$_{10}$) deficiency. *J. Inherit. Metab. Dis.* 38, 145–156 (2014).

Desbats, M. A. *et al.* The *COQ2* genotype predicts the severity of Coenzyme Q_{10} deficiency. *Hum. Mol. Genet.* 25, 4256–4265 (2016).

Di Bella, D. *et al.* Mutations in the mitochondrial protease gene *AFG3L2* cause dominant hereditary ataxia SCA28. *Nat. Genet.* 42, 313–321 (2010).

Di Fonzo, A. *et al. ATP13A2* missense mutations in juvenile parkinsonism and young onset Parkinson disease. *Neurology* 68, 1557–1562 (2007).

Didonna, A. & Opal, P. Advances in Sequencing Technologies for Understanding Hereditary Ataxias: A Review. *JAMA Neurol.* 73, 1485–1490 (2016).

Duina, A., Miller, M. & Keeney, J. Budding Yeast for Budding Geneticists: A Primer on the *Saccharomyces cerevisiae* Model System. *Genetics* 197, 33–48 (2014).

Durr, A. Autosomal dominant cerebellar ataxias: polyglutamine expansions and beyond. *Lancet Neurol.* 9, 885–894 (2010).

Dutta, S., Rieche, F., Eckl, N., Duch, C. & Kretzschmar, D. Glial expression of Swiss cheese (SWS), the Drosophila orthologue of neuropathy target esterase (NTE), is required for neuronal ensheathment and function. *Dis. Model Mech.* 9, 283–294 (2016).

Edenharter, O., Clement, J., Schneuwly, S. & Navarro, J.A. Overexpression of Drosophila frataxin triggers cell death in an iron-dependent manner. *J. Neurogenet.* 31, 189–202 (2017).

Egan, R.J. *et al.* Understanding behavioral and physiological phenotypes of stress and anxiety in zebrafish. *Behav. Brain Res.* 205, 38–44 (2009).

Eiberg, H. *et al.* Novel mutation in *ATP13A2* widens the spectrum of Kufor-Rakeb syndrome (PARK9). *Clin. Genet.* 82, 256–263 (2012).

Emmanuele, V. *et al.* Heterogeneity of Coenzyme Q_{10} Deficiency: Patient Study and Literature Review. *Arch. Neurol.* 69, 978–983 (2012).

Farris, S.M. Are mushroom bodies cerebellum-like structures? *Arthropod Struct Dev*. 40(4):368-79 (2011).

Fitzsimons, H.L. & Scott, M.J. Genetic Modulation of Rpd3 Expression Impairs Long-Term Courtship Memory in Drosophila. *PLoS One* 6, 1–10 (2011).

Fogel, B. *et al.* Exome Sequencing in the Clinical Diagnosis of Sporadic or Familial Cerebellar Ataxia. *JAMA Neurol.* 71, 1237–1246 (2014).

Fogel, B.L. & Perlman, S. Clinical features and molecular genetics of autosomal recessive cerebellar ataxias. *Lancet Neurol.* 6, 245–257 (2007).

Fontana, B.D., Mezzomo, N.J., Kalueff, A.V & Rosemberg, D.B. The developing utility of zebrafish models of neurological and neuropsychiatric disorders: A critical review. *Exp. Neurol.* 299, 157–171 (2018).

Freyer, C. *et al.* Rescue of primary ubiquinone deficiency due to a novel *COQ7* defect using 2,4-dihydroxybensoic acid. *J. Med. Genet.* 52, 779–783 (2015).

Fuchs, S.A. *et al.* Aminoacyl-tRNA synthetase deficiencies in search of common themes. *Genet. Med.* 21, 319–330 (2019).

Gabsi, S. *et al.* Effect of vitamin E supplementation in patients with ataxia with vitamin E deficiency. *Eur. J. Neurol.* 8, 477–481 (2001).

Galatolo, D., Tessa, A., Filla, A. & Santorelli, F.M. 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, 1–8 (2018).

Galatolo, D., *et al*. Bi-allelic mutations in *HARS* cause a multi-system syndrome that includes microcephaly, intellectual, disability, skeletal deformities, and ataxia. Submitted

Galosi, S. *et al*. Dystonia-Ataxia with early handwriting deterioration in *COQ8A* mutation carriers: A case series and literature review. *Parkinsonism Relat. Disord.* 68, 8-16 (2019).

Ganos, C., Hersheson, J., Adams, M., Bhatia, K.P. & Houlden, H. The 4H syndrome due to *RNF216* mutation. *Park. Relat. Disord.* 21, 1122–1123 (2015).

Gazulla, J., Izquierdo-Alvarez, S., Sierra-Martinez, E., Marta-Moreno, M. & Alvarez, S. Inaugural cognitive decline, late disease onset and novel *STUB1* variants in SCAR16. *Neurol. Sci.* 39, 2231–2233 (2018).

Gitler, A.D. *et al.* Alpha-synuclein is part of a diverse and highly conserved interaction network that includes PARK9 and manganese toxicity. *Nat. Genet.* 41, 308–315 (2010).

Goffeau, A. *et al.* Life with 6000 genes. *Science* 274, 563–567 (1996).

Gratz, S.J., Rubinstein, C.D., Harrison, M.M., Wildonger, J. & O'Connor-Giles, K.M. CRISPR-Cas9 Genome Editing in Drosophila. *Curr. Protoc. Mol. Biol.* 111, 1–20 (2015).

Gruenewald, C., Botella, J.A., Bayersdorfer, F., Navarro, J.A. & Schneuwly, S. Hyperoxia-induced neurodegeneration as a tool to identify neuroprotective genes in *Drosophila melanogaster*. *Free Radic. Biol. Med.* 46, 1668–1676 (2009).

Grünewald, A. *et al. ATP13A2* mutations impair mitochondrial function in fibroblasts from patients with Kufor-Rakeb syndrome. *Neurobiol. Aging* 33, 1843.e1-1843.e7 (2012).

Grunwald, D.J. & Eisen, J.S. Headwaters of the zebrafish — emergence of a new model vertebrate. *Nat. Rev. Genet.* 3, 717–724 (2002).

Gusdon, A.M., Zhu, J., Houten, B. Van & Chu, C.T. ATP13A2 regulates mitochondrial bioenergetics through macroautophagy. *Neurobiol. Dis.* 45, 962–972 (2012).

https://www.agilent.com/en/products/cell-analysis/seahorse-xf-consumables/kits-reagents-media/seahorse-xf-cell-mito-stress-test-kit, Agilent Seahorse XF Cell Mito Stress Test Kit

https://droso4schools.wordpress.com/organs/, droso4schools project of Manchester Fly Facility, image of Andreas Prokop

https://medicalxpress.com/news/2015-12-collaborative-brain.html, Hanchuan Peng and HHMI - Janelia Research Campus, image distributed with permission under Vaa3D License, Copyright 2006-2015.

Hales, K.G., Korey, C.A., Larracuente, A.M. & Roberts, D.M. Genetics on the Fly: A Primer on the Drosophila Model System. *Genetics* 201, 815–842 (2015).

Harding, A. The Clinical Features and Classification of the Late Onset Autosomal Dominant Ataxias. *Brain* 105, 1–28 (1982).

Hayer, S.N. *et al. STUB1*/CHIP mutations cause Gordon Holmes syndrome as part of a widespread multisystemic neurodegeneration: evidence from four novel mutations. *Orphanet J. Rare Dis.* 12, 1–8 (2017).

Heidenreich, M. & Zhang, F. Applications of CRISPR–Cas systems in neuroscience. *Nat. Rev. Neurosci.* 17, 36–44 (2016).

Heimdal, K. *et al. STUB1* mutations in autosomal recessive ataxias – evidence for mutation-specific clinical heterogeneity. *Orphanet J. Rare Dis.* 9, 1–12 (2014).

Helbig, K.L. *et al.* A Recurrent Mutation in *KCNA2* as a Novel Cause of Hereditary Spastic Paraplegia and Ataxia. *Ann. Neurol.* 80, 1–5 (2016).

Hewamadduma, C.A. *et al.* Novel genotype-phenotype and MRI correlations in a large cohort of patients with *SPG7* mutations. *Neurol. Genet.* 4, e279 (2018).

Howe, K. *et al.* The zebrafish reference genome sequence and its relationship to the human genome. *Nature* 496, 498–503 (2013).

Hsiao, C.T. *et al*. Mutational analysis of ITPR1 in a Taiwanese cohort with cerebellar ataxias. *PLoS One* 12, e0187503 (2017).

Hufnagel, R.B. *et al.* Neuropathy target esterase impairments cause Oliver–McFarlane and Laurence–Moon syndromes. *J. Med. Genet.* 52, 85–94 (2015).

Hung, M.W. *et al.* From Omics to Drug Metabolism and High Content Screen of Natural Product in Zebrafish: A New Model for Discovery of Neuroactive Compound. *Evidence-based Complement. Altern. Med.* 2012, 1–20 (2012).

Invernizzi, F. *et al.* Mitochondrion Microscale oxygraphy reveals OXPHOS impairment in MRC mutant cells. *Mitochondrion* 12, 328–335 (2012).

Jacob, F., Ho, E.S., Martinez-Ojeda, M., Darras, B.T. & Khwaja, O.S. Case of Infantile Onset Spinocerebellar Ataxia Type 5. *J. Child Neurol.* 28, 1292–1295 (2013).

Jayadev, S. & Bird, T.D. Hereditary ataxias: overview. *Genet. Med.* 15, 673–683 (2013).

Jinek, M. *et al.* A programmable dual RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821 (2012).

Jordanova, A. *et al.* Disrupted function and axonal distribution of mutant tyrosyl-tRNA synthetase in dominant intermediate Charcot-Marie-Tooth neuropathy. *Nat. Genet.* 38, 197–202 (2006).

Kabashi, E., Brustein, E., Champagne, N. & Drapeau, P. Zebrafish models for the functional genomics of neurogenetic disorders. *Biochim. Biophys. Acta* 1812, 335–345 (2011).

Kalueff, A. V *et al.* Towards a Comprehensive Catalog of Zebrafish behaviour 1.0 and beyond. *Zebrafish* 10, 70–86 (2013).

Kawarai, T., Miyamoto, R., Shimatami, Y., Orlacchio, A. & Kaji, R. Choreoathetosis, Dystonia, and Myoclonus in 3 Siblings with Autosomal Recessive Spinocerebellar Ataxia Type 16. *JAMA Neurol.* 73, 888–890 (2016).

Keiser, M.S., Kordasiewicz, H.B. & Mcbride, J.L. Gene suppression strategies for dominantly inherited neurodegenerative diseases: lessons from Huntington's disease and spinocerebellar ataxia. *Hum. Mol. Genet.* 25, R53–R54 (2016).

Khare, S. *et al.* A mutation causes a neurodevelopmental, non-progressive SCA13 subtype associated with dominant negative effects and aberrant EGFR trafficking. *PLoS One* 12, 1–21 (2017).

Kim, A.J., Lazar, A.A. & Slutskiy, Y.B. Projection neurons in Drosophila antennal lobes signal the acceleration of odor concentrations. *Elife* 14, 1–11 (2015).

Kjaergaard, S., Skovby, F. & Schwartz, M. Absence of homozygosity for predominant mutations in *PMM2* in Danish patients with carbohydrate-deficient glycoprotein syndrome type 1. *Eur. J. Hum. Genet.* 6, 331–336 (1998).

Klockgether, T., Mariotti, C. & Paulson, H.L. Spinocerebellar ataxia. *Nat. Rev. Dis. Prim.* 5, 24 (2019).

Kong, S.M.Y. *et al.* Parkinson's disease-linked human PARK9/ATP13A2 maintains zinc homeostasis and promotes Alpha-Synuclein externalization via exosomes. *Hum. Mol. Genet.* 23, 2816–2833 (2014).

Koppen, M., Metodiev, M.D., Casari, G., Rugarli, E.I. & Langer, T. Variable and Tissue-Specific Subunit Composition of Mitochondrial m-AAA Protease Complexes Linked to Hereditary Spastic Paraplegia. *Mol. Cell. Biol.* 27, 758–767 (2007).

Kozol, R. A. *et al.* Function Over Form: Modeling Groups of Inherited Neurological Conditions in Zebrafish. *Front. Mol. Neurosci.* 9, 55 (2016).

Lai, J.S., Lo, S., Dickson, B.J. & Chiang, A. Auditory circuit in the Drosophila brain. *Proc. Natl. Acad. Sci.* 109, 2607–2612 (2012).

Langheinrich, U. Zebrafish: a new model on the pharmaceutical catwalk. *Bioessays* 25, 904–912 (2003).

Laredj, L.N., Licitra, F. & Puccio, H.M. The molecular genetics of coenzyme Q biosynthesis in health and disease. *Biochimie* 100, 78–87 (2014).

Le Bizec, C. *et al.* A New Insight into *PMM2* Mutations in the French Population. *Hum. Mutat.* 25, 504–505 (2005).

Lee, J. *et al.* De Novo Mutations in the Motor Domain of KIF1A Cause Cognitive Impairment, Spastic Paraparesis, Axonal Neuropathy, and Cerebellar Atrophy. *Hum. Mutat.* 36, 69–78 (2015).

Lee, Y. *et al.* Mutations in *KCND3* Cause Spinocerebellar Ataxia Type 22. *Ann. Neurol.* 72, 859–869 (2012).

Lei, L. *et al.* Disorders Safety and efficacy of valproic acid treatment in SCA3/MJD patients. *Park. Relat. Disord.* 26, 55–61 (2016).

Lieto, M. *et al.* The complex phenotype of spinocerebellar ataxia type 48 (SCA48) in eight unrelated Italian families. *Eur. J. Neurol.* (2019).

Lieto, M. *et al.* Overt hypogonadism may be not a sentinel sign of *RNF216*: two novel mutations associated with ataxia, chorea and fertility. *Mov. Disord.* In press (2019).

Lieto, M. *et al*. Progressive myoclonic ataxia associated with ATP13A2 mutation: widening Kufor-Rakeb phenotype? Submitted

Lin, C. H. *et al.* Novel *ATP13A2* variant associated with Parkinson disease in Taiwan and Singapore. *Neurology* 71, 1727–1732 (2008).

Linderman, J.A., Chambers, M.C., Gupta, A.S. & Schneider, D.S. Infection-Related Declines in Chill Coma Recovery and Negative Geotaxis in *Drosophila melanogaster. PLoS One* 7, 1–7 (2012).

Liu, J. & Baraban, S.C. Network Properties Revealed during Multi-Scale Calcium Imaging of Seizure Activity in Zebrafish. *eNeuro* 6, 1–13 (2019).

Lu, M. *et al.* Clinical phenotype, in silico and biomedical analyses, and intervention for an East Asian population-specific c.370G>A (p.G124S) *COQ4* mutation in a Chinese family with CoQ10 deficiency-associated Leigh syndrome. *J. Hum. Genet.* 64, 297–304 (2019).

Macrae, C.A. & Peterson, R.T. Zebrafish as tools for drug discovery. *Nat. Rev. Drug Discov.* 14, 721–731 (2015).

Madrigal, S.C. *et al*. Changes in protein function underlie the disease spectrum in patients with CHIP mutations. *J Biol Chem*. 294(50):19236-19245 (2019).

Magri, S. *et al.* Concurrent *AFG3L2* and *SPG7* mutations associated with syndromic parkinsonism and optic atrophy with aberrant OPA1 processing and mitochondrial network fragmentation. *Hum. Mutat.* 39, 2060–2071 (2018).

Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science* 339, 823–826 (2013).

Malicdan, M.C.V *et al.* A novel inborn error of the coenzyme Q10 biosynthesis pathway: cerebellar ataxia and static encephalomyopathy due to COQ5 C-methyltransferase deficiency. *Hum. Mutat.* 39, 69–79 (2019).

Mancini, C. *et al.* Prevalence and phenotype of the c.1529C>T *SPG7* variant in adult-onset cerebellar ataxia in Italy. *Eur. J. Neurol.* 26, 80–86 (2019).

Mancini, C. *et al.* Mice harbouring a SCA28 patient mutation in *AFG3L2* develop late-onset ataxia associated with enhanced mitochondrial proteotoxicity. *Neurobiol. Dis.* 124, 14–28 (2019).

Marbois, B., Gin, P., Gulmezian, M. & Clarke, C.F. The yeast Coq4 polypeptide organizes a mitochondrial protein complex essential for coenzyme Q biosynthesis. *Biochim. Biophys. Acta* 1791, 69–75 (2009).

Marco, P. De *et al.* GPER signalling in both cancer-associated fibroblasts and breast cancer cells mediates a feedforward IL1β/IL1R1 response. *Sci. Rep.* 6, 1–14 (2016).

Margolin, D. *et al.* Ataxia, Dementia, and Hypogonadotropism Caused by Disordered Ubiquitination. *N. Engl. J. Med.* 368, 1992–2003 (2013).

Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–381 (2005).

Matsumoto, K., Toh-e, A. & Oshima, Y. Genetic Control of Galactokinase Synthesis in *Saccharomyces cerevisiae*: Evidence for Constitutive Expression of the Positive Regulatory Gene gal4. *J. Bacteriol.* 134, 446–457 (1978).

Matthijs, G., Schollen, E., Heykants, L. & Grunewald, S. Phosphomannomutase Deficiency: The Molecular Basis of the Classical Jaeken Syndrome (CDGS Type Ia). *Mol. Genet. Metab.* 68, 220–226 (1999).

Matthijs, G. *et al.* Mutations in *PMM2*, a phosphomannomutase gene on chromosome 16p13, in carbohydrate-deficient glycoprotein type I syndrome (Jaeken syndrome). *Nat. Genet.* 16, 88–92 (1997).

Matthijs, G., Schollen, E., Schaftingen, E. Van, Cassiman, J. & Jaeken, J. Lack of Homozygotes for the Most Frequent Disease Allele in Carbohydrate-Deficient Glycoprotein Syndrome Type 1A. *Am. J. Hum. Genet.* 62, 542–550 (1998).

McGurk, L., Berson, A. & Bonini, N.M. Drosophila as an *In Vivo* Model for Human Neurodegenerative Disease. *Genetics* 201, 377–402 (2015).

McLaughlin, H.M. *et al.* A Recurrent Loss-of-Function Alanyl-tRNA Synthetase (*AARS*) Mutation in Patients with Charcot-Marie-Tooth Disease Type 2N (CMT2N). *Hum. Mutat.* 33, 244–253 (2012).

Meyer-Schuman, R. & Antonellis, A. Emerging mechanisms of aminoacyl-tRNA synthetase mutations in recessive and dominant human disease. *Hum. Mol. Genet.* 26, R114–R127 (2017).

Michele, G. De *et al.* Spinocerebellar ataxia 48 presenting with ataxia associated with cognitive, psychiatric, and extrapyramidal features: A report of two Italian families. *Park. Relat. Disord.* 65, 91–96 (2019).

Mignot, C. *et al.* Phenotypic variability in ARCA2 and identification of a core ataxic phenotype with slow progression. *Orphanet J. Rare Dis.* 8, 173 (2013).

Monin, M. *et al.* 29 French adult patients with PMM2-congenital disorder of glycosylation: outcome of the classical pediatric phenotype and depiction of a late-onset phenotype. *Orphanet J. Rare Dis.* 9, 207 (2014).

Morgan, T.M. Sex Limited Inheritance in Drosophila. *Science* 32, 120–122 (1910).

Morganti, S. *et al.* Complexity of genome sequencing and reporting: Next generation sequencing (NGS) technologies and implementation of precision medicine in real life. *Crit. Rev. Oncol. Hematol.* 133, 171–182 (2019).

Morino, H. *et al.* A mutation in the low voltage-gated calcium channel *CACNA1G* alters the physiological properties of the channel, causing spinocerebellar ataxia. *Mol. Brain* 8, 89 (2015).

Mubaidin, A. & Dasouki, M. Pallido-pyramidal degeneration, supranuclear upgaze paresis and dementia: Kufor-Rakeb syndrome. *Acta Neurol. Scand.* 89, 347–352 (1994).

Nangle, L., Zhang, W., Xie, W., Yang, X. & Schimmel, P. Charcot-Marie-Tooth disease-associated mutant tRNA synthetases linked to altered dimer interface and neurite distribution defect. *Proc. Natl. Acad. Sci.* 104, 11239–11244 (2007).

Ng, S.B. *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nat. Genet.* 42, 30–35 (2010).

Nicholls, D.G. *et al.* Bioenergetic Profile Experiment using C2C12 Myoblast Cells. *J. Vis. Exp.* 1–6 (2010).

Nogi, Y., Shimada, H., Matsuzaky, Y., Hashimoto, H. & Fukasawa, T. Regulation of expression of the galactose gene cluster in Saccharomyces cerevisiae. *Mol. Gen. Genet.* 195, 29–34 (1984).

Norton, W.H.J. Toward developmental models of psychiatric disorders in zebrafish. *Front. Neural Circuits* 7, 79 (2013).

Oprescu, S.N., Griffin, L.B., Beg, A.A. & Antonellis, A. Predicting the pathogenicity of aminoacyl-tRNA synthetase mutations. *Methods* 113, 139–151 (2017).

Orr, H. *et al.* Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nat. Genet.* 4, 221–226 (1993).

Paolacci, S. *et al.* Specific combinations of biallelic *POLR3A* variants cause Wiedemann-Rautenstrauch syndrome. *J. Med. Genet.* 55, 837–846 (2018).

Pareek, G., Thomas, R.E. & Pallanck, L.J. Loss of the Drosophila m-AAA mitochondrial protease paraplegin results in mitochondrial dysfunction, shortened lifespan, and neuronal and muscular degeneration. *Cell Death Dis.* 9, 304 (2018).

Park, J.S., Blair, N.F. & Sue, C.M. The role of ATP13A2 in Parkinson's disease: Clinical phenotypes and molecular mechanisms. *Mov. Disord.* 30, 770–779 (2015).

Parker, M.O., Millington, M.E., Combe, F.J. & Brennan, C.H. Housing Conditions Differentially Affect Physiological and Behavioural Stress Responses of Zebrafish, as well as the Response to Anxiolytics. *PLoS Genet.* 7, 1–9 (2012).

Patron, M., Sprenger, H.G. & Langer, T. m-AAA proteases, mitochondrial calcium homeostasis and neurodegeneration. *Cell Res.* 28, 296–306 (2018).

Pena, S. & Coimbra, R. Ataxia and myoclonic epilepsy due to a heterozygous new mutation in *KCNA2*: proposal for a new channelopathy. *Clin. Genet.* 87, e1–e3 (2015).

Periyasamy, S., Hinds Jr, T., Shemshedini, L., Shou, W. & Sanchez, E. R. FKBP51 and Cyp40 are positive regulators of androgen-dependent prostate cancer cell growth and the targets of FK506 and cyclosporin A. *Oncogene* 29, 1691–1701 (2009).

Pierson, T.M. *et al.* Whole-Exome Sequencing Identifies Homozygous *AFG3L2* Mutations in a Spastic Ataxia-Neuropathy Syndrome Linked to Mitochondrial m-AAA Proteases. *PLoS Genet.* 7, 1–10 (2011).

Prüßing, K., Voigt, A., & Schulz, J.B. Drosophila melanogaster as a model organism for Alzheimer's disease. *Mol Neurodegener.* 22;8:35 (2013).

Puffenberger, E.G. *et al.* Genetic Mapping and Exome Sequencing Identify Variants Associated with Five Novel Diseases. *PLoS One* 7, e28936 (2012).

Qi, L.S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* 152, 1173–1183 (2013).

Qi, Y., Liu, H., Daniels, M. P., Zhang, G. & Xu, H. Loss of Drosophila i-AAA protease, *dYME1L*, causes abnormal mitochondria and apoptotic degeneration. *Cell Death Differ.* 23, 291–302 (2016).

Quelhas, D., Quental, R., Vilarinho, L., Amorim, A. & Azevedo, L. Congenital Disorder of Glycosylation Type Ia: Searching for the Origin of Common Mutations in *PMM2*. *Ann. Hum. Genet.* 71, 348–53 (2006).

Rainier, S. *et al.* Neuropathy Target Esterase Gene Mutations Cause Motor Neuron Disease. *Am. J. Hum. Genet.* 82, 780–785 (2008).

Ramirez, A. *et al.* Hereditary parkinsonism with dementia is caused by mutations in *ATP13A2*, encoding a lysosomal type 5 P-type ATPase. *Nat. Genet.* 38, 1184–1191 (2006).

Ramonet, D. *et al.* PARK9-associated ATP13A2 localizes to intracellular acidic vesicles and regulates cation homeostasis and neuronal integrity. *Hum. Mol. Genet.* 21, 1725–1743 (2012).

Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* 8, 2281–2308 (2013).

Rao, L., Sha, Y. & Eissa, N.T. The E3 ubiquitin ligase *STUB1* regulates autophagy and mitochondrial biogenesis by modulating TFEB activity. *Mol. Cell Oncol.* 4, e1372867 (2017).

Reiter, L.T., Potocki, L., Chien, S., Gribskov, M. & Bier, E. A Systematic Analysis of Human Disease-Associated Gene Sequences In *Drosophila melanogaster*. *Genome Res.* 11, 1114–1125 (2001).

Richards, S. *et al*. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 17(5):405-24 (2015).

Rico, E. P. *et al.* Zebrafish neurotransmitter systems as potential pharmacological and toxicological targets. *Neurotoxicol. Teratol.* 33, 608–617 (2011).

Rohani, M. *et al.* Action Myoclonus and Seizure in Kufor-Rakeb Syndrome. *Mov. Disord. Clin. Pract.* 5, 195–199 (2017).

Romano, S. *et al.* Riluzole in patients with hereditary cerebellar ataxia: a randomised, double-blind, placebo-controlled trial. *Lancet Neurol.* 14, 985–991 (2015).

Rosemberg, D.B. *et al.* Behavioral effects of taurine pretreatment in zebrafish acutely exposed to ethanol. *Neuropharmacology* 63, 613–623 (2012).

Rossi, A. *et al.* Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature* 524, 230–233 (2015).

Royer-Bertrand, B. *et al.* Peripheral neuropathy and cognitive impairment associated with a novel monoallelic *HARS* variant. *Ann. Clin. Transl. Neurol.* 6, 1072–1080 (2019).

Ruano, L. & Silva, M.C. The Global Epidemiology of Hereditary Ataxia and Spastic Paraplegia: A Systematic Review of Prevalence Studies. *Neuroepidemiology* 42, 174–183 (2014).

Sacco, T. *et al.* Mouse brain expression patterns of *Spg7*, *Afg3l1*, and *Afg3l2* transcripts, encoding for the mitochondrial m-AAA protease. *BMC Neurosci.* 11, (2010).

Safka Brozkova, D. *et al.* Loss of function mutations in *HARS* cause a spectrum of inherited peripheral neuropathies. *Brain* 138, 2161–2172 (2015).

Salviati, L. *et al.* Haploinsufficiency of *COQ4* causes coenzyme Q_{10} deficiency. *J. Med. Genet.* 49, 187–191 (2012).

Salviati, L. *et al*. Primary Coenzyme Q$_{10}$ Deficiency. *GeneReviews®* 1993-2019.

Samorodnitsky, E. *et al.* Comparison of Custom Capture for Targeted Next-Generation DNA Sequencing. *J. Mol. Diagnostic* 17, 64–75 (2015).

Sánchez-Ferrero, E. *et al. SPG7* mutational screening in spastic paraplegia patients supports a dominant effect for some mutations and a pathogenic role for p.A510V. *Clin. Genet.* 83, 257–262 (2013).

Sanger, F., Nicklen, S. & Coulson, A. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* 74, 5463–5467 (1977).

Santens, P. *et al. RNF216* mutations as a novel cause of autosomal recessive Huntington-like disorder. *Neurology* 84, 1760–1766 (2015).

Sawyer, S.L. *et al.* Exome Sequencing as a Diagnostic Tool for Pediatric-Onset Ataxia. *Hum. Mutat.* 35, 45–49 (2014).

Schnekenberg, R.P. *et al.* De novo point mutations in patients diagnosed with ataxic cerebral palsy. *Brain* 138, 1817–1832 (2015).

Schniepp, R. *et al.* 4-Aminopyridine and cerebellar gait: a retrospective case series. *J. Neurol.* 259, 2491–2493 (2012).

Schollen, E. *et al.* Increased recurrence risk in congenital disorders of glycosylation type Ia (CDG-Ia) due to a transmission ratio distortion. *J. Med. Genet.* 41, 877–880 (2004).

Schollen, E., Kjaergaard, S., Legius, E., Schwartz, M. & Matthijs, G. Lack of Hardy-Weinberg equilibrium for the most prevalent *PMM2* mutation in CDG-Ia (congenital disorders of glycosylation type Ia). *Eur. J. Hum. Genet.* 8, 367–371 (2000).

Schultheis, P. J. *et al.* Characterization of the P5 subfamily of P-type transport ATPases in mice. *Biochem. Biophys. Res. Commun.* 323, 731–738 (2004).

Scoles, D. R. & Pulst, S. M. Oligonucleotide therapeutics in neurodegenerative diseases. *RNA Biol.* 15, 707–714 (2018).

Sha, Y., Rao, L., Settembre, C., Ballabio, A. & Eissa, N. T. STUB1 regulates TFEB-induced autophagy-lysosome pathway. *EMBO J.* 36, 2544–2552 (2017).

Shi, C. *et al.* Ataxia and hypogonadism caused by the loss of ubiquitin ligase activity of the U box protein CHIP. *Hum. Mol. Genet.* 23, 1013–1024 (2014).

Shi, Y. *et al.* Identification of CHIP as a Novel Causative Gene for Autosomal Recessive Cerebellar Ataxia. *PLoS One* 8, e81884 (2013).

Sondheimer, N. *et al.* Novel recessive mutations in *COQ4* cause severe infantile cardiomyopathy and encephalopathy associated with CoQ_{10} deficiency. *Mol. Genet. Metab. Reports* 12, 23–27 (2017).

Spindler, M., Beal, M.F. & Henchcliffe, C. Coenzyme Q_{10} effects in neurodegenerative disease. *Neuropsychiatr. Dis. Treat.* 5, 597–610 (2009).

Stewart, A.M. *et al.* Molecular psychiatry of zebrafish. *Mol. Psychiatry* 20, 2–17 (2015).

Stewart, A.M., Braubach, O., Spitsbergen, J., Gerlai, R. & Kalueff, A. V. Zebrafish models for translational neuroscience research: from tank to bedside. *Trends Neurosci.* 37, 264–278 (2014).

Sun, M. *et al.* Targeted exome analysis identifies the genetic basis of disease in over 50% of patients with a wide range of ataxia-related phenotypes. *Genet. Med.* 21, 195–206 (2019).

Suster, M.L., Seugnet, L., Bate, M. & Sokolowski, M.B. Refining GAL4-Driven Transgene Expression in Drosophila with a GAL80 Enhancer-Trap. *Genesis* 39, 240–245 (2004).

Svenstrup, K. *et al.* SCA28: Novel Mutation in the AFG3L2 Proteolytic Domain Causes a Mild Cerebellar Syndrome with Selective Type-1 Muscle Fiber Atrophy. *Cerebellum* 16, 62–67 (2017).

Synofzik, M., Puccio, H.M., Mochel, F. & Schols, L. Review Autosomal Recessive Cerebellar Ataxias: Paving the Way toward Targeted Molecular Therapies. *Neuron* 101, 560–583 (2019).

Synofzik, M. *et al.* *PNPLA6* mutations cause Boucher-Neuhauser and Gordon Holmes syndromes as part of a broad neurodegenerative spectrum. *Brain* 137, 69–77 (2014).

Synofzik, M. *et al.* De novo *ITPR1* variants are a recurrent cause of early-onset ataxia, acting via loss of channel function. *Eur. J. Hum. Genet.* 26, 1623–1634 (2018).

Synofzik, M. & Németh, A. Recessive ataxias. *Handb. Clin. Neurol.* 155, 73–89 (2018).

Synofzik, M. & Schüle, R. Overcoming the divide between ataxias and spastic paraplegias: Shared phenotypes, genes, and pathways. *Mov. Disord.* 32, 332–345 (2017).

Synofzik, M. *et al.* Phenotype and frequency of *STUB1* mutations: next-generation screenings in Caucasian ataxia and spastic paraplegia cohorts. *Orphanet J. Rare Dis.* 9, 1–8 (2014).

Synofzik, M. *et al.* SYNE1 ataxia is a common recessive ataxia with major non-cerebellar features: a large multi-centre study. *Brain* 139, 1378–1393 (2016).

Synofzik, M. *et al.* Autosomal recessive spastic ataxia of Charlevoix Saguenay (ARSACS): expanding the genetic, clinical and imaging spectrum. *Orphanet J. Rare Dis.* 8, 1–13 (2013).

Syrbe, S. *et al.* De novo loss- or gain-of-function mutations in *KCNA2* cause epileptic encephalopathy. *Nat. Genet.* 47, 393–399 (2015).

Takahashi, K *et al*. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*. 31(5):861-72 (2007).

Tarnutzer, A.A. *et al*. Boucher-Neuhäuser syndrome: cerebellar degeneration, chorioretinal dystrophy and hypogonadotropic hypogonadism: two novel cases and a review of 40 cases from the literature. *J Neurol*. 262, 194-202 (2015).

Thisse, C. & Thisse, B. High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nat. Protoc.* 3, 59–69 (2008).

Tran, U.C. & Clarke, C.F. Endogenous synthesis of coenzyme Q in eukaryotes. *Mitochondrion* 7S, S62–S71 (2007).

Tulli, S. *et al.* Pathogenic variants in the AFG3L2 proteolytic domain cause SCA28 through haploinsufficiency and proteostatic stress-driven OMA1 activation. *J. Med. Genet.* 56, 499–511 (2019).

Tunc, S. *et al.* Spinocerebellar Ataxia Type 28-Phenotypic and Molecular Characterization of a Family with Heterozygous and Compound-Heterozygous Mutations in *AFG3L2*. *Cerebellum* 18, 817–822 (2019).

Turkgenc, B. *et al. STUB1* polyadenylation signal variant AACAAA does not affect polyadenylation but decreases *STUB1* translation causing. *Hum. Mutat.* 39, 1344–1348 (2018).

Ugur, B., Chen, K. & Bellen, H.J. Drosophila tools and assays for the study of human diseases. *Dis. Model Mech.* 9, 235–244 (2016).

Usenovic, M., Tresse, E., Mazzulli, J.R., Taylor, J.P. & Krainc, D. Deficiency of ATP13A2 Leads to Lysosomal Dysfunction, Alpha-Synuclein Accumulation, and Neurotoxicity. *J. Neurosci.* 32, 4240–4246 (2012).

van de Warrenburg, B.P. *et al.* Clinical exome sequencing for cerebellar ataxia and spastic paraplegia uncovers novel gene–disease associations and unanticipated rare disorders. *Eur. J. Hum. Genet.* 24, 1460–1466 (2016).

van der Windt, G.J.W., Chang, C. & Pearce, E.L. Measuring Bioenergetics in T Cells Using a Seahorse Extracellular Flux Analyzer. *Curr. Protoc. Immunol.* 113, 1–14 (2016).

Van Dijk, T *et al*. A de novo missense mutation in the inositol 1,4,5-triphosphate receptor type 1 gene causing severe pontine and cerebellar hypoplasia: Expanding the phenotype of ITPR1-related spinocerebellar ataxia's. *Am. J. Med. Genet*. 173, 207-212 (2017).

van Gassen, K.L.I. *et al.* Genotype-phenotype correlations in spastic paraplegia type 7: a study in a large Dutch cohort. *Brain* 135, 2994–3004 (2012).

Vaz, R., Hofmeister, W. & Lindstrand, A. Zebrafish Models of Neurodevelopmental Disorders: Limitations and Benefits of Current Tools and Techniques. *Int. J. Mol. Sci.* 20, 1–26 (2019).

Vester, A. *et al.* A Loss-of-Function Variant in the Human Histidyl-tRNA Synthetase (*HARS*) Gene is Neurotoxic *In Vivo*. *Hum. Mutat.* 34, 191–199 (2013).

Vuillaumier-Barrot, S. *et al.* Identification of four novel *PMM2* mutations in congenital disorders of glycosylation (CDG) Ia French patients. *J. Med. Genet.* 37, 579–580 (2000).

Wallen, R.C. & Antonellis, A. To Charge or Not to Charge: Mechanistic Insights into Neuropathy-Associated tRNA Synthetase Mutations. *Curr. Opin. Genet. Dev.* 71, 233–236 (2013).

Wambach, J.A. *et al.* Bi-allelic *POLR3A* Loss-of-Function Variants Cause Autosomal-Recessive Wiedemann-Rautenstrauch Syndrome. *Am. J. Hum. Genet.* 103, 968–975 (2018).

Wan, J. *et al.* Mutations in the RNA exosome component gene *EXOSC3* cause pontocerebellar hypoplasia and spinal motor neuron degeneration. *Nat. Genet.* 44, 704–708 (2012).

Wang, K., Li, M. & Hakonarson, H. Analysing biological pathways in genome-wide association studies. *Nat. Rev. Genet.* 11, 843–854 (2010).

Wang, Y. *et al.* Pathogenicity of two *COQ7* mutations and responses to 2, 4-dihydroxybenzoate bypass treatment. *J. Cell. Mol. Med.* 21, 2329–2343 (2017).

Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet. Genet.* 10, 57–63 (2009).

Warner, J.P. *et al.* A general method for the detection of large CAG repeat expansions by fluorescent PCR. *J. Med. Genet. Genet.* 33, 1022–6 (1996).

Watts, J. K. & Corey, D. R. Silencing disease genes in the laboratory and the clinic. *J. Pathol.* 226, 365–379 (2012).

Wei, Q. *et al.* Regulation of IL-4 receptor signaling by STUB1 in lung inflammation. *Am. J. Respir. Crit. Care Med.* 189, 16–29 (2014).

Williams, B. & Warman, M. CRISPR/Cas9 Technologies. *J. Bone Miner. Res.* 32, 883–888 (2017).

Wu, Y. *et al.* Clinical study and *PLA2G6* mutation screening analysis in Chinese patients with infantile neuroaxonal dystrophy. *Eur. J. Neurol.* 16, 240–245 (2009).

Zambonin, J.L. *et al.* Spinocerebellar ataxia type 29 due to mutations in *ITPR1*: a case series and review of this emerging congenital ataxia. *Orphanet J. Rare Dis.* 12, 121 (2017).

Zanni, G. *et al.* Exome sequencing in a family with intellectual disability, early onset spasticity, and cerebellar atrophy detects a novel mutation in *EXOSC3*. *Neurogenetics* 14, 247–250 (2013).

Zhang, M. *et al.* Chaperoned ubiquitylation--crystal structures of the CHIP U box E3 ubiquitin ligase and a CHIP-Ubc13-Uev1a complex. *Mol. Cell* 20, 525–538 (2005).

Zhang, X. *et al.* Mutations in *QARS*, encoding glutaminyl-tRNA synthetase, cause progressive microcephaly, cerebral-cerebellar atrophy, and intractable seizures. *Am. J. Hum. Genet.* 94, 547–558 (2014).

Zhu, Y. The Drosophila visual system: From neural circuits to behavior. *Cell Adhes. Migr.* 7, 333–344 (2013).

Supplemental material

<u>Full gene lists of multi-gene panels used:</u>

Ataxome (Haloplex, Agilent Technologies): *ABCB7, ACO2, AFG3L2, AMPD2, ANO10, APTX, ATP1A3, ATP2B3, ATP8A2, BEAN1, BRF1, C10ORF2, CA8, CACNA1A, CACNB4, CAMTA1, CASK, CHMP1A, CLP1, COQ2, COQ8A, EEF2, ELOVL4, EXOSC3, FGF14, FLVCR1, GRID2, GRM1, IFRD1, ITPR1, KCNA1, KCNC3, KCND3, KCNJ10, KIAA0226, KIF1A, KIF1C, MARS2, MRE11A, MTPAP, NOP56, NPC1, OPHN1, PDYN, PIK3R5, PLEKHG4, PMPCA, PNKP, PNPLA6, PRKCG, RARS2, RNF216, SACS, SETX, SIL1, SLC1A3, SNX14, SPG7, SPTBN2, STUB1, SYNE1, SYT14, TDP1, TGM6, TPP1, TRPC3, TSEN54, TTBK2, TTC19, TTPA, TUBB4A, UBR4, VAMP1, VLDLR, VRK1, WDR81, WFS1, WWOX, ZFYVE26, ZNF592.*

Ataxome 2.0 (Nimblegen, Roche): *AARS2, ABCB7, ACO2, AFG3L2, AHI1, ALDH5A1, AMPD2, ANO10, APTX, ARL13B, ATCAY, ATG5, ATM, ATP1A3, ATP2B3, ATP8A2, BEAN1, BRAT1, BRF1, C10orf2, C5orf42, CA8, CACNA1A, CACNA1G, CACNB4, CAMTA1, CASK, CC2D2A, CDK5, CEP290, CEP41, CHMP1A, CLN6, CLP1, COQ2, COQ4, COQ8A, COQ9, CSPP1, CSTB, CTBP1, CWF19L1, CYP7B1, DARS, DARS2, DDHD2, DKC1, DNAJC19, DNAJC3, DNMT1, EEF2, EIF2B1, EIF2B2, EIF2B3, EIF2B4, EIF2B5, ELOVL4, EXOSC3, EXOSC8, FARS2, FASTKD2, FGF14, GAN, GBA2, GBE1, GFAP, GJC2, GRID2, GRM1, HARS, HARS2, HEXA, HIBCH, HSD17B4, INPP5E, ITPR1, KCNA1, KCNA2, KCNC3, KCND3, KCNJ10, KCTD7, KIAA0226, KIF1A, KIF1C, KIF7, LAMA1, LMNB2, MARS2, MKS1, MRE11A, MTPAP, MVK, NDUFS1, NOP56, NPC1, NPHP1, OFD1, OPHN1, PAX6, PDE6D, PDHA1, PDSS1, PDSS2, PDYN, PEX10, PEX7, PHYH, PIK3R5, PLA2G6, PLEKHG4, PLP1, PMM2, PMPCA, PNKP, PNPLA6, POLG, POLR3A, POLR3B, PPP2R2B, PRKCG, PRPS1, PRRT2, PSAP, PTF1A, PTRH2, QARS, RAB3GAP1, RARS, RARS2, RELN, RNF170, RNF216, RPGRIP1L, SACS, SAMD9L, SCN1A, SCN2A, SCN8A, SEPSECS, SETX, SIL1, SLC1A3, SLC25A46, SLC2A1, SLC33A1, SLC35A2, SLC52A2, SLC6A19, SNAP25, SNX14, SPG11, SPG7, SPTBN2, STUB1, STXBP1, SYNE1, SYT14, TBP, TCTN1, TCTN3, TDP1, TGM6, THG1L , TINF2, TMEM138, TMEM216, TMEM231, TMEM237, TMEM67, TPP1, TRNT1, TRPC3, TSEN2, TSEN34, TSEN54, TTBK2, TTC19, TTPA, TUBB4A, UBA5, UBR4, UCHL1, VAMP1, VARS2, VLDLR, VRK1, WDR81, WFS1, WWOX, ZFYVE26, ZFYVE27, ZNF423, ZNF592.*

Ataxome 3.0 (SureSelect, Agilent Technologies): *AARS2, ABCB7, ABCD1, ABHD12, ACO2, ADGRG1, AFG3L2, AHDC1, AHI1, ALDH5A1, ALG3, ALG6, AMACR, AMPD2, ANO10, APOB, APTX, ARL13B, ARSA, ATCAY, ATG5, ATL1, ATM, ATP13A2, ATP1A3, ATP2B3, ATP7B, ATP8A2, BEAN1, BRAT1, BRF1, C10ORF2, C12ORF65, C5ORF42, C9ORF72, CA8, CACNA1A,*

CACNA1G, CACNB4, CAMTA1, CASK, CC2D2A, CCDC88C, CD40LG, CDK5, CEP104, CEP290, CEP41, CHMP1A, CLCN2, CLN5, CLN6, CLN8, CLP1, COA7, COQ2, COQ4, COQ8A, COQ9, COX20, CP, CSPP1, CSTB, CTBP1, CTSD, CWF19L1, CYP27A1, CYP7B1, DARS, DARS2, DDHD2, DKC1, DNAJC19, DNAJC3, DNMT1, EEF2, EIF2B1, EIF2B2, EIF2B3, EIF2B4, EIF2B5, ELOVL4, ELOVL5, ERCC4, ERCC8, EXOSC3, EXOSC8, FA2H, FARS2, FASTKD2, FGF14, FLVCR1, FMR1, FOLR1, FXN, GALC, GAN, GBA2, GBE1, GDAP2, GFAP, GJB1, GJC2, GLB1, GOSR2, GRID2, GRM1, HARS, HARS2, HEXA, HEXB, HIBCH, HSD17B4, INPP5E, ITPR1, KCNA1, KCNA2, KCNC1, KCNC3, KCND3, KCNJ10, KCTD7, KIF1A, KIF1C, KIF7, LAMA1, LMNB2, LYST, MARS2, MED17, MFN2, MFSD8, MKS1, MMACHC, MME, MRE11A, MTPAP, MTTP, MVK, NAGLU, NDUFS1, NDUFS7, NEU1, NOL3, NOP56, NPC1, NPC2, NPHP1, OFD1, OPA1, OPA3, OPHN1, OTUD4, PAX6, PDE6D, PDHA1, PDSS1, PDSS2, PDYN, PEX10, PEX16, PEX2, PEX6, PEX7, PHYH, PIK3R5, PLA2G6, PLEKHG4, PLP1, PMM2, PMPCA, PNKP, PNPLA6, POLG, POLR3A, POLR3B, PPT1, PRICKLE1, PRKCG, PRNP, PRPS1, PRRT2, PSAP, PSEN1, PTF1A, PTRH2, QARS, RAB3GAP1, RARS, RARS2, RELN, RNF170, RNF216, RPGRIP1L, RUBCN, SACS, SAMD9L, SCN1A, SCN2A, SCN8A, SCYL1, SEPSECS, SETX, SIL1, SLC17A5, SLC1A3, SLC25A46, SLC2A1, SLC33A1, SLC35A2, SLC52A2, SLC6A19, SLC9A1, SLC9A6, SMPD1, SNAP25, SNX14, SPAST, SPG11, SPG7, SPTBN2, SRD5A3, STS, STUB1, STXBP1, SURF1, SYNE1, SYNE2, SYT14, TBC1D23, TCTN1, TCTN2, TCTN3, TDP1, TGM6, THG1L, TINF2, TMEM138, TMEM216, TMEM231, TMEM237, TMEM240, TMEM67, TOP1, TPP1, TRNT1, TRPC3, TSEN2, TSEN34, TSEN54, TTBK2, TTC19, TTC21B, TTPA, TUBB3, TUBB4A, UBA5, UBR4, UCHL1, VAMP1, VARS2, VLDLR, VRK1, VWA3B, WDR81, WFS1, WWOX, XPA, XRCC4, ZFYVE26, ZFYVE27, ZNF423, ZNF592.

Ataxome 4.0 (SureSelect, Agilent Technologies): *AARS2, ABCB7, ABCD1, ABHD12, ACO2, ADGRG1, AFG3L2, AHDC1, AHI1, ALDH5A1, ALG3, ALG6, AMACR, AMPD2, ANO10, APOB, APTX, ARL13B, ARSA, ATCAY, ATG5, ATL1, ATM, ATP13A2, ATP1A2, ATP1A3, ATP2B3, ATP7B, ATP8A2, BEAN1, BRAT1, BRF1, C10ORF2, C12ORF65, C19ORF12, C5ORF42, C9ORF72, CA8, CACNA1A, CACNA1G, CACNB4, CAMTA1, CASK, CC2D2A, CCDC88C, CD40LG, CDK5, CEP104, CEP290, CEP41, CHMP1A, CLCN2, CLN5, CLN6, CLN8, CLP1, COA7, COQ2, COQ4, COQ8A, COQ9, COX20, CP, CSPP1, CSTB, CTBP1, CTSD, CWF19L1, CYP27A1, CYP7B1, DAB1, DARS, DARS2, DDHD2, DKC1, DNAJC19, DNAJC3, DNMT1, EEF2, EIF2B1, EIF2B2, EIF2B3, EIF2B4, EIF2B5, ELOVL4, ELOVL5, ERCC4, ERCC8, EXOSC3, EXOSC8, FA2H, FARS2, FASTKD2, FAT2, FGF14, FLVCR1, FMR1, FOLR1, FXN, GALC, GAN, GBA2, GBE1, GFAP, GJB1, GJC2, GLB1, GOSR2, GRID2, GRM1, HARS, HARS2, HEXA, HEXB, HIBCH, HSD17B4, INPP5E, ITPR1, KCNA1, KCNA2, KCNC1, KCNC3, KCND3, KCNJ10,*

KCNMA1, KCTD7, KIF1A, KIF1C, KIF7, LAMA1, LMNB2, LYST, MARS2, MED17, MFN2, MFSD8, MKS1, MMACHC, MME, MRE11A, MTPAP, MTTP, MVK, NAGLU, NDUFS1, NDUFS7, NEU1, NOL3, NOP56, NPC1, NPC2, NPHP1, OFD1, OPA1, OPA3, OPHN1, PAX6, PCLO, PCNA, PDE6D, PDHA1, PDSS1, PDSS2, PDYN, PEX10, PEX16, PEX2, PEX6, PEX7, PHYH, PIK3R5, PLA2G6, PLD3, PLP1, PMM2, PMPCA, PNKP, PNPLA6, POLG, POLR3A, POLR3B, PPT1, PRICKLE1, PRKCG, PRNP, PRPS1, PRRT2, PSAP, PSEN1, PTF1A, PTRH2, PUM1, QARS, RAB3GAP1, RARS, RARS2, RELN, RNF170, RNF216, RPGRIP1L, RUBCN, SACS, SAMD9L, SARS, SCN1A, SCN2A, SCN8A, SCYL1, SEPSECS, SETX, SIL1, SLC17A5, SLC1A3, SLC25A46, SLC2A1, SLC33A1, SLC35A2, SLC52A2, SLC6A19, SLC9A1, SLC9A6, SMPD1, SNAP25, SNX14, SPAST, SPG11, SPG7, SPTAN1, SPTBN2, SRD5A3, STS, STUB1, STXBP1, SURF1, SYNE1, SYNE2, SYT14, TBC1D23, TCTN1, TCTN2, TCTN3, TDP1, TDP2, TGM6, TINF2, TMEM138, TMEM216, TMEM231, TMEM237, TMEM240, TMEM67, TOE1, TOP1, TPP1, TRNT1, TRPC3, TSEN15, TSEN2, TSEN34, TSEN54, TTBK2, TTC19, TTC21B, TTPA, TUBB3, TUBB4A, UBA5, UBR4, UCHL1, VAMP1, VARS2, VLDLR, VPS53, VRK1, VWA3B, WDR73, WDR81, WFS1, WWOX, XPA, XRCC1, XRCC4, ZFYVE26, ZFYVE27, ZNF423.

Table S1. Primers used in this study (Primers used for Sanger sequencing are not included).

Primer name	Sequence (from 5' to 3')	Application
HARS-RT-Fw	GACAGCCGGGATGATCGAG	RT PCR in skin fibroblasts
HARS-RT-Rv	CTTCTCGCGAACTGCCATCT	RT PCR in skin fibroblasts
COQ4-RT-Fw	ATGAGGAGGGATCCAGAGGG	RT PCR in skin fibroblasts
COQ4-RT-Rv	GGGAGACCCTGTTCACATCC	RT PCR in skin fibroblasts
18S-RT-Fw	GGCGTCCCCCAACTTCTTA	RT PCR in skin fibroblasts
18S-RT-Rv	GGGCATCACAGACCTGTTATTG	RT PCR in skin fibroblasts
SPG7-RT-Fw1	AGCATGAACAGCCCGAGAAG	RT PCR in fruit fly
SPG7-RT-Rv1	CTCTGCTGCTCTTATCGCCAT	RT PCR in fruit fly
SPG7-RT-Fw2	ACTTCTAGCAGTACCATTGGTGA	RT PCR in fruit fly
SPG7-RT-Rv2	GAACGCATCTTTTCGTCATTCTC	RT PCR in fruit fly
AFG3L2-RT-Fw1	AACAAGAAGTGGGTGAGAGTGC	RT PCR in fruit fly
AFG3L2-RT-Rv1	GACACTGCCGATGTTGAACC	RT PCR in fruit fly
AFG3L2-RT-Fw2	ACATCGGCAGTGTCGATAGC	RT PCR in fruit fly
AFG3L2-RT-Rv2	AGTTGATGGATTCCGTTCCCT	RT PCR in fruit fly
RP49-Fw1	CCAAGCACTTCATCCGCCACC	RT PCR in fruit fly
RP49-Rv1	GCGGGTGCGCTTGTTCGATCC	RT PCR in fruit fly
*dSPG7*gRNA1Fw	CTTCGAAAATGTAACAAGTATGGTA	gRNA1 for *dSPG7* KO
*dSPG7*gRNA1Rv	AAACTACCATACTTGTTACATTTTC	gRNA1 for *dSPG7* KO
*dSPG7*gRNA2Fw	CTTCGAAATCTAAGAAAACGTATAA	gRNA2 for *dSPG7* KO
*dSPG7*gRNA2Rv	AAACTTATACGTTTTCTTAGATTTC	gRNA2 for *dSPG7* KO
*dSPG7*hom1AgeIFw	GGCCACCGGTCCACGCGCTATGTGTCGTGG	Homology fragment 1 for *dSPG7* KO
*dSPG7*hom1SacIIRv	GGCCCCGCGGGTATGGTATGGTAATGTC	Homology fragment 1 for *dSPG7* KO
*dSPG7*hom2KpnIFw	GCGCGGTACCTACGTTTTCTTAGATTTTGTGATTTTCC	Homology fragment 2 for *dSPG7* KO
*dSPG7*hom2SpeIRv	GCGCACTAGTGTAAAATAGAGTTCATAGTTGCGC	Homology fragment 2 for *dSPG7* KO

Primer name	Sequence (from 5' to 3')	Application
*dSPG7*KIgRNA1Fw	CTTCGCTGGTCGAGATGGACGGTA	gRNA1 for *dSPG7* KI
*dSPG7*KIgRNA1Rv	AAACTACCGTCCATCTCGACCAGC	gRNA1 for *dSPG7* KI
*dSPG7*KIgRNA2Fw	CTTCGCTTGCCCAGATCGTATGGCG	gRNA2 for *dSPG7* KI
*dSPG7*KIgRNA2Rv	AAACCGCCATACGATCTGGGCAAGC	gRNA2 for *dSPG7* KI
Ori_AmpFw	ACTGGCCGTCGTTTTACAAC	p.Ala510Val vector for *dSPG7* KI
Ori_AmpRv	ATGAGTGAGCTAACTCACATTAATTG	p.Ala510Val vector for *dSPG7* KI
*dSPG7*Hom1KIFw	GTTGTAAAACGACGGCCAGTGAGCCCGCAAAAGGTGTCC	p.Ala510Val vector for *dSPG7* KI
*dSPG7*Hom1KIRv	GCACGCCCTCCTTGGTGGCCATCCCGTCCATCTCGACCAGCAGCTGATTAAGCGTCTGC	p.Ala510Val vector for *dSPG7* KI
*dSPG7*Hom2KIFw	CATACGATCTGGGCAAGCGGCAGGTA	p.Ala510Val vector for *dSPG7* KI
*dSPG7*Hom2KIRv	ATGTGAGTTAGCTCACTCATCGCTAAAGTAAGTAACTTGCCGGTATGGA	p.Ala510Val vector for *dSPG7* KI
RescueKIFw	ATGGCCACCAAGGAGGGCGTGC	p.Ala510Val vector for *dSPG7* KI
RescueKIRv	CCGCTTGCCCAGATCGTATGGGGGCGGTCCGATTAGCTGCACCACCTG	p.Ala510Val vector for *dSPG7* KI
Ala510ValFw	ACGCTTGGTCCGCCTGACACCGGGTTTTTC	p.Ala510Val vector for *dSPG7* KI
Ala510ValRv	GTGTCAGGCGGACCAAGCGTTGTGAGAAG	p.Ala510Val vector for *dSPG7* KI
*coq4*KOgRNAFw	TAATACGACTCACTATAGGGTTACCGGGTGCTTTCCTGTTTTAGAGCTAGAAATAGC	gRNA for *coq4* KO
*coq4*dCas9gRNAFw	TAATACGACTCACTATAGGTGCTTTCCTTGGTTCTTTGTTTTAGAGCTAGAAATAGC	gRNA for *coq4* d-Cas9
coq4-RT-Fw	TGGAAGATTGTATCCAAGCCACA	RT PCR in zebrafish
coq4-RT-Rv	ACCTTCAGGGTCATTTCTCATCC	RT PCR in zebrafish

Table S2. *In silico predictions, frequencies and references of pathogenic mutations found in this study (Legend after the table).*

Index case	Gene (Ref_Seq)	cDNA variant, protein variant	Zigosity	SIFT	Poly Phen-2	Mutation Taster	FATHMM-MKL	Meta SVM	MetaR	GERP	LRT	Mutation Assessor	Provean	CADD score	Frequency (homozygous count)	References
Pt1	*AARS2* (NM_020745)	c.446G>A p.Cys149Tyr	Het	D	D	D	D	D	D	D	D	MD	D	32.0	-	-
Pt1	*AARS2* (NM_020745)	c.385A>C p.Thr129Pro	Het	D	D	D	D	D	D	D	D	HD	D	28.4	$4.06 \times 10^{-6}(0)$	-
Pt2	*AFG3L2* (NM_006796)	c.2167G>A p.Val723Met	Het	D	D	D	D	D	D	D	D	LD	D	32.0	$1.79 \times 10^{-4}(0)$	Tunc et al. 2019
Pt2	*AFG3L2* (NM_006796)	c.634dupG p.Val212Glyfs*4	Het	-	-	-	-	-	-	-	-	-	-	-	-	-
Pt3	*AFG3L2* (NM_006796)	c.2105G>A p.Arg702Gln	Het	D	D	D	D	D	D	D	D	MD	D	33.0	$3.98 \times 10^{-6}(0)$	Di Bella et al. 2010
Pt4	*ANO10* (NM_018075)	c.1009T>G p.Phe337Val	Het	D	D	D	D	T	T	T	D	MD	D	26.9	$4.06 \times 10^{-6}(0)$	-
Pt4	*ANO10* (NM_018075)	c.289delA p.Met97*	Het	-	-	-	-	-	-	-	-	-	-	-	$2.44 \times 10^{-5}(0)$	-
Pt5	*ANO10* (NM_018075)	c.206T>A p.Leu69*	Hom	-	-	-	-	-	-	-	-	-	-	-	-	-
Pt6	*ATM* (NM_000051)	c.2152dupT p.Cys718Leufs*20	Het	-	-	-	-	-	-	-	-	-	-	-	-	-
Pt6	*ATM* (NM_000051)	c.2929T>C p.Cys977Arg	Het	D	D	D	D	T	T	D	D	MD	D	25.9	-	-
Pt7	*ATP13A2* (NM_022089)	c.1205C>T p.Thr402Met	Hom	D	D	D	D	D	D	D	D	HD	D	26.7	$2.85 \times 10^{-5}(0)$	-
Pt8	*CACNA1A* (NM_023035)	c.4897G>A p.Asp1633Asn	Het	D	D	D	D	D	D	D	D	HD	D	34.0	-	-
Pt9	*CACNA1G* (NM_018896)	c.5144G>A p.Arg1715His	Het	D	D	D	D	D	D	D	D	HD	D	35.0	-	Coutelier et al. 2015, Morino et al. 2015
Pt10	*COQ4* (NM_016035)	c.577C>T p.Pro193Ser	Het	D	D	D	D	D	D	D	D	HD	D	33.0	-	-
Pt10	*COQ4* (NM_016035)	c.718C>T p.Arg240Cys	Het	D	D	D	D	T	T	D	D	MD	D	35.0	$1.90 \times 10^{-4}(0)$	Brea-Calvo et al. 2015, Chung et al. 2015
Pt11	*COQ4* (NM_016035)	c.284G>A p.Gly95DAsp	Het	D	D	D	D	D	D	T	N	HD	D	32.0	$4.01 \times 10^{-6}(0)$	-
Pt11	*COQ4* (NM_016035)	c.305G>A p.Arg102His	Het	D	D	D	D	D	T	D	D	HD	D	34.0	-	-
Pt12	*COQ8A* (NM_020247)	c.589-3C>G	Het	-	-	-	-	-	-	-	-	-	-	-	-	Galosi et al, 2019
Pt12	*COQ8A* (NM_020247)	c.1844G>A p.Gly615Asp	Het	D	D	D	D	D	D	D	D	HD	D	33.0	-	Galosi et al, 2019 Salviati et al. GeneReviews® 1993-2019

Index case	Gene (Ref_Seq)	cDNA variant, protein variant	Zigosity	SIFT	Poly Phen-2	Mutation Taster	FATHMM-MKL	Meta SVM	MetaR	GERP	LRT	Mutation Assessor	Provean	CADD score	Frequency (homozygous count)	References
Pt13	COQ8A (NM_020247)	c.1042C>T p.Arg348*	Hom	-	-	-	-	-	-	-	-	-	-	-	4.16×10^{-5}(0)	Galosi et al, 2019
Pt14	EXOSC3 (NM_016042)	c.395A>C p.Asp132Ala	Hom	D	PD	D	D	D	D	D	D	MD	D	28.1	3.82×10^{-4}(0)	Wan et al. 2012, Zanni et al. 2013
Pt15	GJC2 (NM_020435)	c.219_220delCC p.Leu74fValfs*33	Het	-	-	-	-	-	-	-	-	-	-	-	-	-
Pt15	GJC2 (NM_020435)	c.254T>C p.Val85Ala	Het	D	PD	D	D	D	D	T	D	LD	D	23.5	-	-
Pt16	HARS (NM_001289094)	c.616G>T p.Asp206Tyr	Het	D	D	D	D	T	T	T	D	HD	D	31.0	-	-
Pt16	HARS (NM_001289094)	c.730delG p.Val244Cysfs*6	Het	-	-	-	-	-	-	-	-	-	-	-	-	-
Pt17	HARS (NM_001289094)	c.1393A>C p.Ile465Leu	Het	T	D	D	D	D	D	D	D	LD	N	26.6	1.22×10^{-5}(0)	-
Pt17	HARS (NM_001289094)	c.910_912dupTTG p.Leu305dup	Het	-	-	-	-	-	-	-	-	-	-	-	-	-
Pt18	HSD17B4 (NM_001199291)	c.2191C>T p.Gln731*	Het	-	-	-	-	-	-	-	-	-	-	-	1.22×10^{-5}(0)	-
Pt18	HSD17B4 (NM_001199291)	c.727G>T p.Val243Leu	Het	D	B	D	D	D	D	D	D	MD	N	27.5	-	-
Pt19	ITPR1 (NM_001168272)	c.722G>A p.Arg241Lys	Het	D	D	D	D	D	D	D	D	MD	D	29.4	-	Barresi et al., 2016
Pt20	ITPR1 (NM_001168272)	c.805C>T p.Arg269Trp	Het	D	D	D	D	D	D	T	D	MD	D	23.6	-	Barresi et al. 2016, Zambonin et al. 2017, Synofzik et al. 2018
Pt21	ITPR1 (NM_001168272)	c.7748>C p.Ile2583Thr	Het	D	D	D	D	D	D	D	D	HD	D	28.0	-	Van Dijk et al. 2017, Zambonin et al. 2017, Hsiao et al. 2017
Pt22	KCNA2 (NM_004974)	c.890G>A p.Arg297Gln	Het	D	D	D	D	D	D	D	D	HD	D	32.0	-	Pena and Coimbra et al. 2015, Syrbe et al. 2015
Pt23	KCNA2 (NM_004974)	c.881G>A p.Arg294His	Het	D	D	D	D	D	D	D	D	MD	D	27.9	-	Helbig et al. 2016
Pt24	KCNC3 (NM_004977)	c.1268G>A p.Arg423His	Het	D	D	D	D	D	D	T	U	MD	D	33.0	-	Khare et al. 2017
Pt25	KCND3 (NM_004980)	c.680_682delTCT p.Phe227del	Het	-	-	-	-	-	-	-	-	-	-	-	-	Lee et al. 2012
Pt26	KIF1C (NM_006612)	c.765delC p.Asp256Thrfs*10	Hom	-	-	-	-	-	-	-	-	-	-	-	-	-
Pt27	LAMA1 (NM_005559)	c.1404_1405delAG p.Gly469Alafs*5	Het	-	-	-	-	-	-	-	-	-	-	-	-	-
Pt27	LAMA1 (NM_005559)	c.184C>T p.Arg62*	Het	-	-	-	-	-	-	-	-	-	-	-	4.01×10^{-6}(0)	-
Pt28	PLA2G6 (NM_001349864)	c.1703T>C p.Phe568Ser	Het	D	D	D	D	D	D	D	D	MD	D	28.8	-	-

Index case	Gene (Ref_Seq)	cDNA variant, protein variant	Zigosity	SIFT	Poly Phen-2	Mutation Taster	FATHMM-MKL	Meta SVM	MetaR	GERP	LRT	Mutation Assessor	Provean	CADD score	Frequency (homozygous count)	References
Pt28	PLA2G6 (NM_001349864)	c.1111G>A p.Val371Met	Het	D	D	D	D	T	T	D	D	LD	N	34.0	4.06×10^{-6}(0)	Wu et al. 2009
Pt29	PMM2 (NM_000303)	c.422G>A p.Arg141His	Het	D	PD	D	D	D	D	D	D	HD	D	34.0	4.08×10^{-3}(0)	Matthijs et al. 1997, Kjaergaard et al. 1998, Matthijs et al. 1999, Vuillaumier-Barrot et al. 2000, Bohles et al. 2001, Schollen et al. 2006, Quelhas et al. 2006
Pt29	PMM2 (NM_000303)	c.323C>T p.Ala108Val	Het	T	B	D	D	D	D	D	D	MD	D	26.0	1.21×10^{-5}(0)	Matthijs et al. 1998, Schollen et al. 2004, Le Bizec et al. 2005, Monin et al. 2014
Pt30	PNPLA6 (NM_001166111)	c.1880C>T p.Ala627Val	Hom	D	B	D	D	T	T	T	N	MD	D	25.7	8.24×10^{-6}(0)	-
Pt31	PNPLA6 (NM_001166111)	c.2264A>C p.Gln755Pro	Het	D	D	D	D	T	T	D	D	MD	D	27.5	-	-
Pt31	PNPLA6 (NM_001166111)	c.3388C>T p.His1130Tyr	Het	D	D	D	D	T	D	T	D	MD	D	23.6	-	-
Pt32	PNPLA6 (NM_001166111)	c.3023A>G p.Asp1008Gly	Het	D	PD	D	D	D	D	T	D	HD	D	32.0	-	-
Pt32	PNPLA6 (NM_001166111)	c.4075C>T p.Arg1359Trp	Het	D	PD	P	N	T	T	T	N	LD	D	23.5	3.88×10^{-5}(0)	Tarnutzer et al. 2015
Pt33	PNPLA6 (NM_001166111)	c.3365C>T p.Pro1122Leu	Het	D	D	D	D	D	D	T	U	HD	D	27.8	4.07×10^{-6}(0)	Synofzik et al. 2014
Pt33	PNPLA6 (NM_001166111)	c.4081C>T p.Arg1361*	Het	-	-	-	-	-	-	-	-	-	-	-	3.87×10^{-5}(0)	-
Pt34	PNPLA6 (NM_001166111)	c.3385G>A p.Gly1129Arg	Hom	D	D	D	D	D	D	T	D	HD	D	27.0	3.20×10^{-5}(1)	Hufnagel et al. 2015
Pt35	POLR3A (NM_007055)	c.4073G>A p.Gly1358Glu	Het	D	D	D	D	D	D	D	D	HD	D	34.0	-	-
Pt35	POLR3A (NM_007055)	c.1909+22G>A	Het		-	-	-	-	-	-	-	-	-	-	1.34×10^{-3}(0)	Paolacci et al. 2018, Wambach et al. 2018
Pt36	POLR3A (NM_007055)	c.3839dupT p.Met1280Ilefs*20	Het	-	-	-	-	-	-	-	-	-	-	-	-	-
Pt36	POLR3A (NM_007055)	c.1909+22G>A	Het	-	-	-	-	-	-	-	-	-	-	-	1.34×10^{-3}(0)	Paolacci et al. 2018, Wambach et al. 2018
Pt37	PRKCG (NM_002739)	c.358C>T p.Leu120Phe	Het	D	D	D	D	D	D	T	NA	MD	D	25.4	-	-
Pt38	RARS2 (NM_001318785)	c.1037C>T p.Thr346Ile	Het	T	B	D	D	T	T	D	D	LD	N	24.6	-	-

Index case	Gene (Ref_Seq)	cDNA variant, protein variant	Zigosity	SIFT	Poly Phen-2	Mutation Taster	FATHMM-MKL	Meta SVM	MetaR	GERP	LRT	Mutation Assessor	Provean	CADD score	Frequency (homozygous count)	References
Pt38	*RARS2* (NM_001318785)	c.517G>A p.Asp173Asn	Het	D	D	D	D	D	D	D	D	HD	D	34.0	-	-
Pt39	*RNF216* (NM_207111)	c.2061+3A>G	Het	-	-	-	-	-	-	-	-	-	-	-	-	Lieto et al, 2019
Pt39	*RNF216* (NM_207111)	c.1849A>G p.Met617Val	Het	D	B	D	D	T	T	T	D	MD	D	24.2	8.16×10^{-6}(0)	Lieto et al, 2019
Pt40	*SETX* (NM_001351527)	c.3242T>C p.Phe1081Ser	Hom	T	D	D	D	D	D	D	N	LD	D	30.0	-	-
Pt41	*SETX* (NM_001351527)	c.7292dupA p.Asn2431Lysfs*19	Hom	-	-	-	-	-	-	-	-	-	-	-	-	-
Pt42	*SLC2A1* (NM_006516)	c.136C>T p.Gln46*	Het	-	-	-	-	-	-	-	-	-	-	-	-	-
Pt43	*SPG7* (NM_003119)	c.1231G>A p.Asp411Asn	Het	D	D	D	D	D	D	D	D	MD	D	32.0	-	-
Pt43	*SPG7* (NM_003119)	c.679C>T p.Arg227*	Het	-	-	-	-	-	-	-	-	-	-	-	-	-
Pt44	*SPG7* (NM_003119)	c.1529C>T p.Ala510Val	Het	D	D	D	D	D	D	D	D	MD	D	32.0	2.88×10^{-3}(2)	Sanchez-Ferrero et al. 2013, Mancini et al. 2019, Coarelli et al. 2019
Pt44	*SPG7* (NM_003119)	c.1940C>A p.Ala647Glu	Het	D	D	D	D	D	D	D	D	HD	D	25.4	4.08×10^{-6}(0)	-
Pt45	*SPG7* (NM_003119)	c.1529C>T p.Ala510Val	Hom	D	D	D	D	D	D	D	D	MD	D	32.0	2.88×10^{-3}(2)	Sanchez-Ferrero et al. 2013, Mancini et al. 2019, Coarelli et al. 2019
Pt46	*SPG7* (NM_003119)	c.1529C>T p.Ala510Val	Hom	D	D	D	D	D	D	D	D	MD	D	32.0	2.88x10-3(2)	Sanchez-Ferrero et al. 2013, Mancini et al. 2019, Coarelli et al. 2019
Pt47	*SPG7* (NM_003119)	Del exon 2	Hom	-	-	-	-	-	-	-	-	-	-	-	-	-
Pt48	*SPG7* (NM_003119)	Del exon 2	Hom	-	-	-	-	-	-	-	-	-	-	-	-	-
Pt49	*SPG7* (NM_003119)	c.73_80delCCAGGCCC p.Pro25Glyfs*46	Het	-	-	-	-	-	-	-	-	-	-	-	-	-
Pt49	*SPG7* (NM_003119)	c.1940C>A p.Ala647Glu	Het	D	D	D	D	D	D	D	D	HD	D	25.4	4.08×10^{-6}(0)	-
Pt50	*SPG7* (NM_003119)	Del exon 2	Hom	-	-	-	-	-	-	-	-	-	-	-	-	-
Pt51	*SPG7* (NM_003119)	c.1529C>T p.Ala510Val	Het	D	D	D	D	D	D	D	D	MD	D	32.0	2.88×10^{-3}(2)	Sanchez-Ferrero et al. 2013, Mancini et al. 2019, Coarelli et al. 2019

119

Index case	Gene (Ref_Seq)	cDNA variant, protein variant	Zigosity	SIFT	Poly Phen-2	Mutation Taster	FATHMM-MKL	Meta SVM	MetaR	GERP	LRT	Mutation Assessor	Provean	CADD score	Frequency (homozygous count)	References
Pt51	*SPG7* (NM_003119)	c.1972G>A p.Ala658Thr	Het	T	D	D	D	D	D	D	D	HD	D	26.8	-	-
Pt52	*SPTBN2* (NM_006946)	c.1438C>T p.Arg480Trp	Het	D	D	D	D	D	T	T	D	HD	D	33.0	-	Jacob et al. 2012, Parolin Schnekenberg et al. 2015
Pt53	*SPTBN2* (NM_006946)	c.157+1G>A	Het	-	-	-	-	-	-	-	-	-	-	-	-	-
Pt53	*SPTBN2* (NM_006946)	c.1843C>T p.Arg615Trp	Het	D	D	D	D	T	T	T	D	HD	D	24.1	$1.09 \times 10^{-5}(0)$	-
Pt54	*STUB1* (NM_005861)	c.97G>A p.Gly33Ser	Het	D	D	D	D	D	D	T	U	HD	D	24.7	-	De Michele et al. 2019
Pt55	*STUB1* (NM_005861)	c.689_692delACCT p.Tyr230Cysfs*9	Het	-	-	-	-	-	-	-	-	-	-	-	$8.14 \times 10^{-6}(0)$	Depondt et al. 2014, Lieto et al. 2019
Pt56	*STUB1* (NM_005861)	c.682C>T p.Pro228Ser	Het	D	D	D	D	T	T	T	N	HD	D	26.0	-	De Michele et al. 2019
Pt57	*STUB1* (NM_005861)	c.199G>A p.Ala67Thr	Het	D	D	D	D	D	D	T	D	HD	D	29.2	-	Lieto et al. 2019
Pt58	*STUB1* (NM_005861)	c.673C>T p.Arg225*	Het	-	-	-	-	-	-	-	-	-	-	-	$4.08 \times 10^{-6}(0)$	Lieto et al. 2019
Pt59	*STUB1* (NM_005861)	c.721C>T p.Arg241Trp	Het	D	D	D	D	T	T	T	D	MD	D	31.0	$8.15 \times 10^{-6}(0)$	Lieto et al. 2019
Pt60	*STUB1* (NM_005861)	c.170C>T p.Pro57Leu	Het	D	D	D	D	D	D	T	D	MD	D	29.5	-	Lieto et al. 2019
Pt61	*SYNE1* (NM_182961)	c.15049C>T p.Gln5017*	Hom	-	-	-	-	-	-	-	-	-	-	-	-	-
Pt62	*SYNE1* (NM_182961)	c.7085dupA p.Asn2369Lysfs*4	Het	-	-	-	-	-	-	-	-	-	-	-	-	-
Pt62	*SYNE1* (NM_182961)	c.6724-1G>A	Het	-	-	-	-	-	-	-	-	-	-	-	-	-
Pt63	*SYNE1* (NM_182961)	c.4609C>T p.Arg1537*	Hom	-	-	-	-	-	-	-	-	-	-	-	-	-
Pt64	*TMEM216* (NM_001173991)	c.432-1G>C	Hom	-	-	-	-	-	-	-	-	-	-	-	$1.12 \times 10^{-3}(0)$	-
Pt65	*TTPA* (NM_000370)	c.553-1G>T	Hom	-	-	-	-	-	-	-	-	-	-	-	-	-

Table S2 legend:

Zigosity: Het = Heterozygous, Hom = Homozygous

SIFT: D = Damaging, T = Tolerated

Polyphen-2: D = Probably damaging, PD = Possibly damaging, B = Benign

MutationTaster: D = Disease causing, P = Polymorphism

FATHMM-MKL: D = Damaging, N = Neutral

MetaSVM: D = Damaging, T = Tolerated

MetaR: D = Damaging, T = Tolerated

GERP: D = Damaging, T = Tolerated, NA = Not available (GERP RS score: Deleterious threshold >4.4, Dong et al. 2015. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum. Mol. Genet. 24, 2125-2137, 2015)

LRT: D= Deleterious, N = Neutral, U = Unknown

Mutation assessor: HD = High damage, MD = Medium damage, LD = Low damage

Provean: D = Damaging, N = Neutral

NA: Not available

Table S3. In silico predictions, ACMG classifications, frequencies and references of all missense mutations so far described in STUB1 (SCAR16+SCA48). Mutations found in this study are labelled in red. ACMG classifications were reported from the original publication whenever available, otherwise were obtained based on the examination of the corresponding publication (table continues in the next page).

Mutation	CADD	ACMG classification	SIFT	Polyphen-2	Mutation Taster	FATHMM-MKL	MetaSVM	MetalR	GERP	LRT	Mutation assessor	Provean	gnomAD frequency (homozygous count)	References
c.97G>A, p.Gly33Ser	24.7	Likely pathogenic	Damaging	Probably damaging	Disease causing	Damaging	Damaging	Damaging	Tolerated	Unknown	High damage	Damaging	Absent	De Michele et al., 2019
c.103C>A, p.Arg35Ser	23.7	Likely pathogenic	Damaging	Probably damaging	Disease causing	Damaging	Tolerated	Tolerated	Tolerated	Unknown	Low damage	Damaging	Absent	Gazulla et al., 2018
c.170C>T, p.Pro57Leu	29.5	Likely pathogenic	Damaging	Possibly damaging	Disease causing	Damaging	Damaging	Damaging	Tolerated	Deleterious	Medium damaging	Damaging	Absent	Lieto et al., 2019
c.194A>G, p.Asn65Ser	25.8	Pathogenic	Damaging	Probably damaging	Disease causing	Damaging	Damaging	Damaging	Damaging	Deleterious	High damage	Damaging	4.78×10^{-6} (0)	Heimdal et al., 2014
c.199G>A, p.Ala67Thr	29.0	Likely pathogenic	Damaging	Probably damaging	Disease causing	Damaging	Damaging	Damaging	Tolerated	Deleterious	High damage	Damaging	Absent	Lieto et al., 2019
c.235G>A, p.Ala79Thr	30.0	Likely pathogenic	Tolerated	Probably damaging	Disease causing	Damaging	Damaging	Damaging	Tolerated	Deleterious	Medium damage	Neutral	4.06×10^{-6} (0)	Synofzik et al., 2014
c.236C>A, p.Ala79Asp	32.0	Likely pathogenic	Damaging	Probably damaging	Disease causing	Damaging	Damaging	Damaging	Tolerated	Deleterious	High damage	Damaging	Absent	Synofzik et al., 2014
c.367C>G, p.Leu123Val	25.4	Likely pathogenic	Damaging	Probably damaging	Disease causing	Damaging	Damaging	Damaging	Tolerated	Deleterious	Medium damage	Damaging	Absent	Synofzik et al., 2014
c.389A>T, p.Asn130Ile	25.6	Likely pathogenic	Damaging	Probably damaging	Disease causing	Damaging	Tolerated	Tolerated	Damaging	Deleterious	Medium damage	Damaging	Absent	Shi et al., 2013
c.433A>C, p.Lys145Gln	23.9	Pathogenic	Tolerated	Benign	Disease causing	Damaging	Tolerated	Tolerated	Damaging	Deleterious	Medium damage	Neutral	6.01×10^{-4} (1)	Coutelier et al., 2017 Depondt et al., 2014 Hayer et al., 2017 Sun et al., 2019
c.441G>T, p.Trp147Cys	32.0	Likely pathogenic	Damaging	Probably damaging	Disease causing	Damaging	Tolerated	Tolerated	Damaging	Deleterious	Medium damage	Damaging	Absent	Shi et al., 2013
c.493C>T, p.Leu165Phe	26.5	Pathogenic	Damaging	Probably damaging	Disease causing	Damaging	Tolerated	Tolerated	Damaging	Deleterious	Medium damage	Damaging	Absent	Shi et al., 2013
c.502C>T, p.Leu168Phe	25.9	Pathogenic	Damaging	Probably damaging	Disease causing	Damaging	Tolerated	Tolerated	Damaging	Deleterious	Medium damage	Damaging	Absent	Coutelier et al., 2017
c.633G>A, p.Met211Ile	22.6	Pathogenic	Tolerated	Benign	Disease causing	Damaging	Tolerated	Tolerated	Damaging	Deleterious	Low damage	Neutral	Absent	Bettencourt et al., 2015

Mutation	CADD	ACMG classification	SIFT	Polyphen-2	Mutation Taster	FATHMM-MKL	MetaSVM	MetalR	GERP	LRT	Mutation assessor	Provean	gnomAD frequency (homozygous count)	References
c.682C>T, p.Pro228Ser	26.0	Likely pathogenic	Damaging	Probably damaging	Disease causing	Damaging	Tolerated	Tolerated	Tolerated	Neutral	High damage	Damaging	Absent	De Michele et al., 2019
c.707G>C, p.Ser236Thr	25.9	Pathogenic	Tolerated	Probably damaging	Disease causing	Damaging	Tolerated	Tolerated	Tolerated	Deleterious	Low damage	Neutral	Absent	Shi et al., 2013
c.719T>C, p.Met240Thr	25.5	Likely pathogenic	Damaging	Probably damaging	Disease causing	Damaging	Tolerated	Tolerated	Tolerated	Deleterious	High damage	Damaging	Absent	Synofzik et al., 2014
c.721C>G, p.Arg241Gly	24.9	Likely pathogenic	Damaging	Probably damaging	Disease causing	Damaging	Tolerated	Tolerated	Tolerated	Deleterious	Medium damage	Damaging	4.00×10^{-6} (0)	Sun et al., 2019
c.721C>T, p.Arg241Trp	31.0	Likely pathogenic	Damaging	Probably damaging	Disease causing	Tolerated	Tolerated	Tolerated	Tolerated	Deleterious	Medium damaging	Damaging	7.99×10^{-6} (0)	Lieto et al., 2019
c.724G>A, p.Glu242Lys	26.1	Likely pathogenic	Damaging	Probably damaging	Disease causing	Damaging	Tolerated	Tolerated	Tolerated	Deleterious	Medium damage	Damaging	Absent	Kawarai et al., 2016
c.728C>T, p.Pro243Leu	25.2	Likely pathogenic	Damaging	Probably damaging	Disease causing	Damaging	Damaging	Damaging	Tolerated	Deleterious	High damage	Damaging	4.00×10^{-6} (0)	Hayer et al., 2017
c.737C>T, p.Thr246Met	24.9	Pathogenic	Damaging	Probably damaging	Disease causing	Damaging	Tolerated	Tolerated	Tolerated	Deleterious	High damage	Damaging	0 (0)	Shi et al., 2014
c.823C>G, p.Leu275Val	23.7	Likely pathogenic	Damaging	Probably damaging	Disease causing	Damaging	Tolerated	Tolerated	Tolerated	Deleterious	Medium damage	Damaging	Absent	Cordoba et al., 2014
c.880A>T, p.Ile294Phe	24.7	Likely pathogenic	Damaging	Probably damaging	Disease causing	Damaging	Tolerated	Tolerated	Tolerated	Neutral	Medium damage	Damaging	Absent	Hayer et al., 2017

Table S4. ACMG classifications, frequencies and references of all non-missense mutations so far described in STUB1 (SCAR16+SCA48). Mutations found in this study are labelled in red. Variants that we found in parallel with this NGS study (p.Val264Glyfs*4, p.Pro274Alafs*3, p.Leu275Aspfs*16), not listed among our positive cases, but published together (Lieto et al., 2019) were included in the table and labelled in red as well. ACMG classifications were reported from the original publication whenever available, otherwise were obtained based on the examination of the corresponding publication.

Mutation	ACMG classification	gnomAD frequency (homozygous count)	References
c.355C>T, p.Arg119*	Pathogenic	Absent	Hayer et al., 2017
c.430A>T, p.Lys144*	Pathogenic	Absent	Heimdal et al., 2014
c.612+1G>C	Pathogenic	Absent	Cordoba et al., 2014
c.621C>G, p.Tyr207*	Pathogenic	Absent	Shi et al., 2013
c.646dupT, p.Ser216Phefs*5	Pathogenic	Absent	Sun et al., 2019
c.673C>T, p.Arg225*	Pathogenic	7.109×10^{-6} (0)	Lieto et al., 2019
c.678_679delCA, p.Ile227Phefs*11	Pathogenic	Absent	Gazulla et al., 2018
c.689_692delACCT, p.Tyr230Cysfs*9	Pathogenic	7.985×10^{-6} (0)	Lieto et al., 2019 Depondt et al., 2014
c.694_699delTGTGGC, p.Cys232_Gly233del	Likely pathogenic	Absent	Sun et al., 2019
c.712G>T, p.Glu238*	Pathogenic	Absent	Bettencourt et al., 2015
c.791_792delTG, p.Val264Glyfs*4	Pathogenic	Absent	Lieto et al., 2019
c.818_819dupGC, p.Pro274Alafs*3	Pathogenic	Absent	Lieto et al., 2019
c.823_824delCT, p.Leu275Aspfs*16	Pathogenic	3.99×10^{-6} (0)	Genis et al., 2018 Lieto et al., 2019
c*240T>C	Likely pathogenic	Absent	Turkgenc et al., 2018