



# The History of Gene Hunting in Hereditary Spinocerebellar Degeneration: Lessons From the Past and Future Perspectives

Ashraf Yahia<sup>1,2,3,4</sup> and Giovanni Stevanin<sup>3,4\*</sup>

<sup>1</sup>Department of Biochemistry, Faculty of Medicine, University of Khartoum, Khartoum, Sudan, <sup>2</sup>Department of Biochemistry, Faculty of Medicine, National University, Khartoum, Sudan, <sup>3</sup>Institut du Cerveau, INSERM U1127, CNRS UMR7225, Sorbonne Université, Paris, France, <sup>4</sup>Ecole Pratique des Hautes Etudes, EPHE, PSL Research University, Paris, France

## OPEN ACCESS

### Edited by:

Ayman Hussein,  
University of Khartoum, Sudan

### Reviewed by:

Zhipeng Liu,  
Purdue University, United States  
Bolin Chen,  
Northwestern Polytechnical  
University, China

### \*Correspondence:

Giovanni Stevanin  
giovanni-b.stevanin@inserm.fr

### Specialty section:

This article was submitted to  
Human and Medical Genomics,  
a section of the journal  
Frontiers in Genetics

**Received:** 07 December 2020

**Accepted:** 02 March 2021

**Published:** 23 March 2021

### Citation:

Yahia A and Stevanin G (2021)  
The History of Gene Hunting in  
Hereditary Spinocerebellar  
Degeneration: Lessons From the  
Past and Future Perspectives.  
Front. Genet. 12:638730.  
doi: 10.3389/fgene.2021.638730

Hereditary spinocerebellar degeneration (SCD) encompasses an expanding list of rare diseases with a broad clinical and genetic heterogeneity, complicating their diagnosis and management in daily clinical practice. Correct diagnosis is a pillar for precision medicine, a branch of medicine that promises to flourish with the progressive improvements in studying the human genome. Discovering the genes causing novel Mendelian phenotypes contributes to precision medicine by diagnosing subsets of patients with previously undiagnosed conditions, guiding the management of these patients and their families, and enabling the discovery of more causes of Mendelian diseases. This new knowledge provides insight into the biological processes involved in health and disease, including the more common complex disorders. This review discusses the evolution of the clinical and genetic approaches used to diagnose hereditary SCD and the potential of new tools for future discoveries.

**Keywords:** hereditary spastic paraplegia, hereditary cerebellar ataxia, spinocerebellar ataxia, spinocerebellar degeneration, gene discovery, diagnosis, neurogenetics

## INTRODUCTION

Hereditary forms of spastic paraplegia (SPG), cerebellar ataxia (CA), spastic ataxia, and spinocerebellar ataxia (SCA) are distinct clinical entities caused by related mechanisms and encompassing a continuum of phenotypes known as hereditary spinocerebellar degenerations (SCDs; Parodi et al., 2018). To date, they have been shown to be caused by pathogenic variants in more than 200 genes and loci inherited through X-linked, autosomal recessive, autosomal dominant, and mitochondrial patterns, and occasionally due to *de novo* events (Synofzik and Schüle, 2017; Parodi et al., 2018; Boutry et al., 2019; Manto et al., 2020; see **Supplementary Table 1**). SCDs are characterized clinically by cerebellar ataxia and/or spastic limbs, often complicated by other neurological or extra-neurological features (Parodi et al., 2018). The genetic and phenotypic heterogeneity of SCDs complicates their diagnosis (Elsayed et al., 2019). Moreover, the overlap between the clinical presentations of SCD and those of other genetic and non-genetic neurological conditions further complicates diagnosing these diseases (Elsayed et al., 2019). These diagnostic hurdles require the use of constantly evolving tools and approaches.

Spinocerebellar degenerations are estimated to affect 1:10,000 individuals worldwide, and thus fall within the category of rare diseases (Ruano et al., 2014; Nguengang Wakap et al., 2020). More than 70% of rare diseases are genetic, the majority of which affect the nervous system (Lee et al., 2020; Nguengang Wakap et al., 2020). From a biological perspective, Mendelian diseases can be viewed as natural “knockouts” or gains of function of a gene, a pathway, or a cellular process that can provide insight into complex biological systems and the more common complex diseases (Lee et al., 2020). From epidemiological and clinical perspectives, Mendelian diseases affect >2.5% of the human population (Nguengang Wakap et al., 2020), and therefore reaching a clinical diagnosis is valuable for the patients affected by these conditions and their families (Clark et al., 2018; Wright et al., 2018). The discovery of the cause of a Mendelian disease has, on some occasions, had a snowball effect and led to the discovery of more diseases (van der Knaap et al., 2002; Abou Jamra et al., 2011). Finding more causative genes helps to complete the puzzle and increases the probability of identifying biomarkers and therapeutic targets, hopefully of interest for multiple clinical-genetic entities.

The identification of SCD genes has passed several milestones, and multiple tools have been used in this quest (Figure 1; Supplementary Table 1). The evolution of the number of SCD entities has greatly benefited from improvements in terms of the tools used, their cost, the clinical phenotyping, and the availability of multiple samples in affected families. Several reviews have addressed the best practices to diagnose SCD subtypes at different periods (Harding, 1993; Fink, 1997; Brusse et al., 2007; De Silva et al., 2019; Shribman et al., 2019). Here, we review the evolution of the approaches used to identify

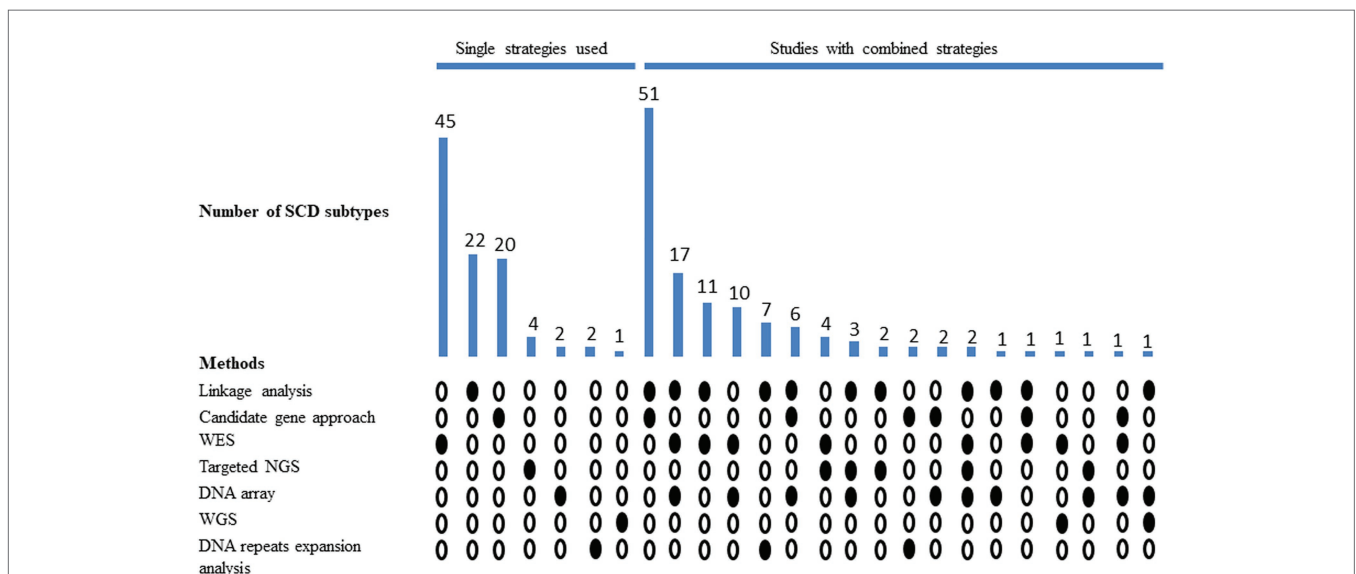
novel SCD genes over time and discuss the possibility of combining some of these approaches with novel tools to increase the current SCD diagnostic rates and answer the current challenges in these diseases.

## TOOLS USED TO DISCOVER SCD GENES IN THE PAST

### Linkage Analysis and Homozygosity Mapping

Whole genome linkage analyses, and homozygosity mapping in consanguineous cases, were the primary approaches for identifying genes and loci associated with Mendelian diseases (Lipner and Greenberg, 2018). These methods were developed in the 1980s thanks to improvements in detection of genetic markers and their assignment in chromosomal maps. These strategies were and are still, suitable for family-based approaches and realizable only in large pedigrees, although single nuclear sibships can be used in autosomal recessive consanguineous families (Lipner and Greenberg, 2018).

The linkage analysis concept relies on genetic variant co-segregation with nearby markers at the disease locus, as recombination is less likely to separate adjacent loci (Lipner and Greenberg, 2018). Microsatellite markers and array-based single nucleotide polymorphism (SNP) markers have been extensively used in SCD gene discovery. These markers replaced the restriction-fragment-length polymorphisms (RFLP) and the variable numbers of tandem repeats (VNTR) used in the 1980s. SNP markers are now commonly used either with DNA microarray or extracted from exome/genome sequencing data (Guergueltcheva et al., 2012; Novarino et al., 2014).



**FIGURE 1 |** Methods used in diagnosing novel subtypes of hereditary spinocerebellar degeneration (SCD). Filled circles represent the methods used in identifying SCD subtypes, and the bar chart shows the number of SCD subtypes identified by each method or combination of methods. Empty circles represent the unused methods. SCD, hereditary spinocerebellar degeneration; WES, whole-exome sequencing; NGS, next-generation sequencing; and WGS, whole-genome sequencing.

Among the 56% (123/220) of known SCD subtypes identified by approaches that involved linkage analysis, a good example of its diagnostic utility is the study by Novarino et al. (2014) that reported 15 novel causative genes in consanguineous families with various forms of SPG; they used SNP marker-based homozygosity to filter and prioritize the exome data. Indeed, most of the linkage-based SCD studies identified the responsible genes by coupling linkage analysis with candidate genes screening (51/220) or, more recently through biased (2/218) or unbiased next-generation sequencing (NGS; 31/220). For example, Guerguelcheva et al. (2012) pinpointed the mutation responsible for autosomal recessive spinocerebellar ataxia (SCAR) type 13 (SCAR13; OMIM # 614831) using linkage studies and mutational screening. There are still 22 SCD loci for which the causative gene has not been found and may have been missed if the causative variant is in non-coding regions, often remaining unexplored. Indeed, chromosomal location through linkage studies was critical for the identification of several of the nucleotide repeat expansion disorders such as the recently reported “cerebellar ataxia neuropathy and vestibular areflexia” syndrome (CANVAS) due to intronic repeat expansions in *RFC1* (Cortese et al., 2019).

The use of microsatellite or SNP markers is not confined to identifying disease loci but can extend to establishing the founder effects of disease-causing variants (Engert et al., 2000; Zivony-Elboum et al., 2012). This is clearly illustrated in studies performed to determine the ancestral origin of the pathological expansions in SCA3/Machado-Joseph disease (OMIM # 109150; Martins et al., 2007; Sharony et al., 2019).

## Repeat Expansion Detection Methods

Expansion of simple DNA sequence repeats causes more than 40 diseases, most with phenotypes affecting the central nervous system (Paulson, 2018; Rodriguez and Todd, 2019). They were first identified in a series of SCA loci in the 1990s following linkage studies and DNA/cDNA cloning strategies. These included the studies performed by Koide et al. (1994), David et al. (1997), and Koob et al. (1999) that directly screened patients with an autosomal dominant pattern of inheritance for repeat expansions and identified the genes responsible for dentatorubral-pallidoluysian atrophy (DRPLA; OMIM # 125370), SCA7 (OMIM # 164500), and SCA8 (OMIM # 608768). These discoveries benefited from the development of repeat expansion detection methods that used oligonucleotide probes containing triplet repeats and antibodies against polyglutamine tracks (Li et al., 1993; Trottier et al., 1995; Bahlo et al., 2018; Mitsuhashi and Matsumoto, 2020). The 1990s also witnessed the identification of the cause of Friedreich ataxia (OMIM # 229300), the most frequent autosomal recessive cerebellar ataxia commonly caused by intronic repeat expansions in the *FXN* gene (Campuzano et al., 1996).

There has been renewed interest in repeat expansion in recent years, with the development of diagnostic applications of PCR-based approaches (Cagnoli et al., 2018) and the improvement of algorithms for short-read sequence analysis. Recently, a combination of linkage analysis and short-read whole-genome sequencing (WGS) identified repeat expansions in *DABI* and *RFC1* causing SCA37 (OMIM # 615945) and

CANVAS (OMIM # 614575), respectively (Seixas et al., 2017; Cortese et al., 2019). In all likelihood, long-read sequencing methods will soon further increase the power of detecting repeat expansions (Bahlo et al., 2018; De Roeck et al., 2019; Mitsuhashi and Matsumoto, 2020). Up to now, methods for detecting DNA repeat expansions were applied in discovering 5% (11/220) of the SCD subtypes, mostly (9/11) when coupled with whole genome linkage analysis.

## Candidate Gene Approaches

Prioritizing genes is a complementary but crucial step in all the approaches taken to identify SCD genes. This prioritization usually depends on the gene expression, the gene's function or its product's function, and the phenotype of patients and/or animal models. Interestingly, a straightforward candidate gene approach was the main methodology in identifying 9.1% (20/220) of SCD subtypes.

There are multiple mouse models of neurodegeneration, particularly with cerebellar features, and they can be very informative in guiding candidate gene selection. Escayg et al. (2000) screened patients with neurological phenotypes for mutations in the *CACNB4* gene after they had identified bi-allelic mutations in *CACNB4* in a mouse model presenting with ataxia and epilepsy, and they identified *CACNB4* as the cause of type 5 episodic ataxia (EA5; OMIM # 613855; Burgess et al., 1997). Loss of function variants in the *NHE1* gene coding the sodium-hydrogen exchanger 1 (NHE1) are associated with Lichtenstein-Knorr syndrome (OMIM# 616291), a cerebellar syndrome complicated by sensorineural hearing loss (Guissart et al., 2015). Interestingly, the Lichtenstein-Knorr syndrome's phenotype is partially observed in the mouse knockout for *NHE1* and for *CHP1* as well, coding an NHE1 interactor protein (Bell et al., 1999; Liu et al., 2013). These observations prompted Mendoza-Ferreira et al. (2018) to prioritize a frameshift variant in *CHP1* observed on whole-exome sequencing (WES) data of Moroccan patients with complex cerebellar ataxia and intellectual disability, the first cases of spastic ataxia type 9 (OMIM # 618438). Human genetics can also benefit animal genetics. This is shown by the identification of a *KIFC* variant in Charolais cattle developing a spastic ataxia phenotype by prioritization of the WGS data on this gene after it was involved in a similar human disorder (Duchesne et al., 2018).

The clinical and biological features of the patients have also helped to point to specific candidate genes. The patients' clinical presentation and enzymatic assays significant for ubiquinone deficiency led to the implication of pathogenic mutations in *ADCK3* in SCAR type 9 (SCAR9; OMIM # 612016; Mollet et al., 2008). The patients' phenotype and hormonal assays implying thyroid hormone resistance led to the implication of *SLC16A2* mutations in X-linked SPG22 (OMIM # 300523; Friesema et al., 2004). Similarly, the clinical phenotype led to the discovery of some SCD subtypes caused by mitochondrial alterations (Holt et al., 1990; Winterthun et al., 2005; Sarzi et al., 2007).

Many attempts to identify the SCD genes' interactors have been performed using yeast two-hybrid screening (Lim et al., 2006) and *in silico* networks (Novarino et al., 2014). The latter study identified three causative genes in HSP

families using an HSP *in silico* functional network (Novarino et al., 2014). Based on this “HSP network,” the authors selected three candidate genes and tested them by direct sequencing in a large series of HSP cases, a strategy that led to *REEP2*, *MAG*, and *BICD2* being identified as novel HSP genes (Novarino et al., 2014). Along the same lines, Ślabicki et al. (2010) adopted another approach in identifying the gene responsible for SPG48; they first performed RNAi genome-wide screening in HeLa cells for genes involved in DNA double-strand breaks and identified the *AP5Z1* gene among the modifier genes. *AP5Z1* was shown to interact with spastascin and spastizin, the proteins involved in SPG11 and SPG15, respectively. Ślabicki et al. (2010) then sequenced the *AP5Z1* gene in 166 patients with unexplained HSP and identified the first SPG48 case. These studies demonstrated the usefulness of knowing the interactors of the SCD proteins as they are likely the source of potential variants in patients.

Knowledge of the physiopathology at the cell level can also guide the candidate gene analysis. Jen et al. (2005) screened a cohort of patients with unexplained episodic ataxia and hemiplegic migraine for mutations in *SLC1A3*, which encodes a glutamate transporter essential for removing glutamate from the synaptic cleft. Glutamate is the most abundant amino acid neurotransmitter in the central nervous system. In their study, Jen et al. (2005) identified a heterozygous mutation in *SLC1A3* in a patient with episodic ataxia called later episodic ataxia (EA) type 6 (EA6; OMIM # 612656).

Other studies have implicated genes known for their association with specific neurological phenotypes in developing new SCD forms based on phenotypic similarities between the old and the new phenotypes. These include studies that implicated *SCN2A* in EA9 (OMIM # 618924; Liao et al., 2010), *GJC2* in SPG44 (OMIM# 613206; Orthmann-Murphy et al., 2009), *PGN* in the autosomal dominant SPG7 (Sánchez-Ferrero et al., 2013), *FMR1* in fragile X tremor/ataxia syndrome (FXTAS, OMIM # 300623; Hagerman et al., 2001), and *DARS2* in hereditary SPG (Lan et al., 2017).

## DNA Microarrays

DNA microarrays are used in clinical settings to investigate copy number variations (CNVs) and can also help determine SNP genotypes for linkage analysis or homozygosity mapping strategies (Levy and Burnside, 2019). DNA microarrays used for CNV detection are the first-tier tests for diagnosing neurodevelopmental disorders, disease entities that include intellectual disabilities, developmental delay disorders, autism spectrum disorders, and disorders with multiple congenital anomalies (Jang et al., 2019). However, a recent meta-analysis suggested that WES outperforms microarrays in diagnosing neurodevelopmental disorders (Srivastava et al., 2019). Furthermore, in a recent study performed over a 10-year period, Ciaccio et al. (2020) discouraged the use of chromosomal microarrays as the first-tier test for diagnosing neurodevelopmental disorders with ataxia.

DNA arrays were the sole genetic screening tool used in discovering two SCD genes. Moreno-De-Luca et al. (2011) used a custom-designed oligonucleotide array in discovering *AP4E1*, the gene responsible for SPG51 (OMIM# 613744).

Moreno-De-Luca et al. (2011) used a genome-wide microarray in their study to identify a region of interest, and then further analyzed the identified region with a higher-resolution microarray. In another study, Utine et al. (2013) used an SNP array in discovering the gene causing SCAR18 (OMIM # 616204). Guided by array-based homozygosity mapping, Utine et al. (2013) identified deletions in the third and fourth exons of *GRID2* in three Turkish siblings. They confirmed the absence of these exons in the patients using real-time PCR (RT-PCR).

Coupling arrays to other genetic screening tools has led to the identification of 20% (44/220) of all currently known SCD genes. More than half of the genes thus identified (27/44) were discovered using approaches that involved microarrays (mainly for linkage studies) and WES.

## Next-Generation Sequencing Approaches

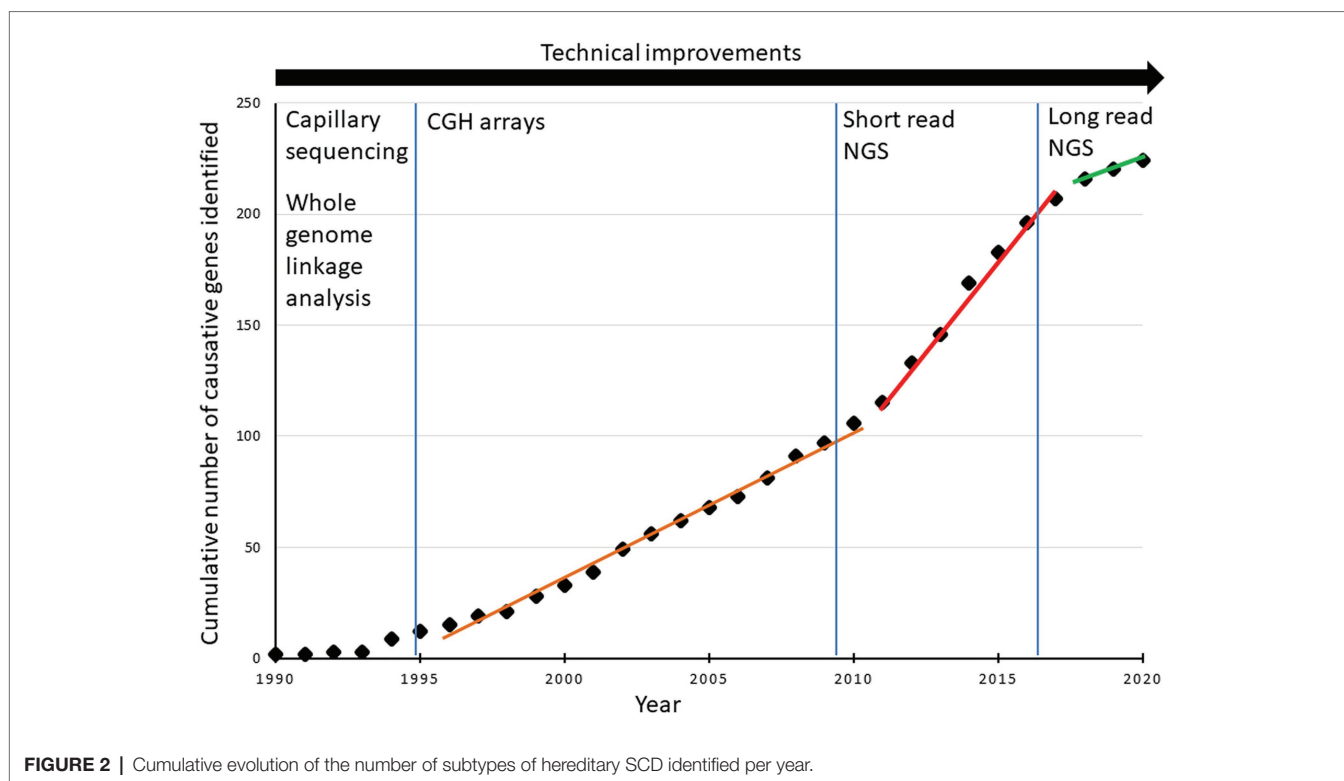
The advent of NGS has revolutionized the field of medical genetics as it enabled parallel sequencing of massive targets in a short duration of time and for a plummeting cost (Boycott et al., 2017; Mazzarotto et al., 2020). NGS is classified according to the targeted sequences. The targeted sequence in WGS is the whole genome, in WES it is the coding part of the genome, and in targeted NGS (TS) it is a custom-ordered set of genes and sequences (Gao and Smith, 2020). NGS is also classified as long-read sequencing and short-read sequencing according to the length of the sequenced fragments, also known as the sequence reads. The length of the sequence read in short-read sequencing is ~75–300 base-pairs, while in long-read sequencing; it is equal to the length of the template sequence, at least in theory (Caspar et al., 2018).

Next-generation sequencing is extensively and routinely used in diagnosing Mendelian diseases, including those with neurological phenotypes (Bhatia et al., 2020; Rouleau et al., 2020). To date, 47.3% (104/220) of SCD genes have been discovered using NGS-based approaches, and the slope of the curve in SCD gene discovery drastically changed from 2009, when this technique was made available (Figure 2).

## Targeted NGS

Targeted NGS is the most commonly used NGS tool in clinical practice (Gao and Smith, 2020). In the clinical context, TS investigates the roles of a predetermined set of genes in a clinical phenotype (Gao and Smith, 2020). In the setting of research and gene discovery, researchers use TS to focus on specific regions and sequences of interest determined by the clinical data or other diagnostic modalities.

Targeted NGS was involved in the discovery of 7.3% (16/220) of known SCD subtypes, mostly (73%) when coupled with other diagnostic modalities. TS was the primary genetic test used in four studies that identified four new SCD subtypes. In the first, Zanni et al. (2012) sequenced all the exons in the X chromosome in a family with X-linked cerebellar ataxia and identified a missense mutation in the *ATP2B3* gene. Pathogenic mutations in *ATP2B3* cause X-linked SCA type 1 (SCAX1; OMIM # 302500; Zanni et al., 2012). In the other three studies, TS identified mutations in *SPTBN2*, *KIF1A*, and



*TUBB4A* as the causes of SCAR14 (OMIM # 615386), the autosomal dominant form of SPG30 (OMIM # 610357), and *TUBB4A*-associated SPG, respectively (Lise et al., 2012; Ylikallio et al., 2015; Sagnelli et al., 2016). Pathogenic mutations in *SPTBN2*, *KIF1A*, and *TUBB4A* were already known to cause SCA5 (OMIM # 600224), autosomal recessive SPG30 (OMIM # 610357), and both autosomal dominant type 4 dystonia (OMIM # 128101) and type 6 hypomyelination leukodystrophy (OMIM # 612438), respectively (Ikeda et al., 2006; Erlich et al., 2011; Hersheson et al., 2013; Simons et al., 2013). Thus, the three subsequent studies added SCD to the clinical phenotypes associated with genes previously known to cause diseases other than SCD, a situation becoming more frequent in recent years than real gene identification in SCD as shown by the change in the slope of the curve in **Figure 2** since 2017.

Targeted NGS is mainly used in genetic diagnosis in clinical practice as it allows focused analysis, thereby limiting incidental findings (Elsayed et al., 2016; Morais et al., 2017). In addition, given that it is optimized for a specific set of genes, sequence capture is usually homogeneous, allowing CNV detection through coverage analysis (Moreno-Cabrera et al., 2020).

### Whole-Exome Sequencing

Whole-exome sequencing covers ~1–2% of the human genome and is extensively used for diagnosing Mendelian diseases (Hansen et al., 2020; Rouleau et al., 2020). The 3' and 5' untranslated regions (UTRs) are also included in most WES kits, at a variable level according to the company; however, variants in these regions are often difficult to analyze and interpret (Devanna et al., 2018). To date, mutations in ~20%

of the protein-coding genes are known associated with a disease trait (Posey, 2019).

There are two views on whether TS or WES is the best first-tier test for diagnosing patients with SCDs (Galatolo et al., 2018; Shribman et al., 2019). TS has a lower cost, higher depth, lower analysis time, and fewer incidental findings (Gao and Smith, 2020; Mazzarotto et al., 2020; Platt et al., 2020). WES has higher coverage and outperforms TS in discovering new Mendelian disease-causing genes (Gao and Smith, 2020; Mazzarotto et al., 2020; Platt et al., 2020). WES was the only genetic screening tool used in discovering 20.5% (45/220) of known SCD subtypes and an additional 21.4% (47/220) when coupled with other techniques. Focused exome sequencing (FES) is a WES-based approach that possesses some of the advantages of TS and WES (Pengelly et al., 2020). FES has higher coverage than TS, as it currently targets ~5,000 disease-associated genes, ~20% of the sequences targeted by WES (Pengelly et al., 2020). FES has greater depth and produces fewer incidental findings compared to WES but has a lower potential for discovering new SCD genes (Pengelly et al., 2020).

The diagnostic yield of WES is higher in familial cases compared to sporadic cases (Coutelier et al., 2018; Ngo et al., 2020). However, this generalization has some exceptions. For example, Fogel et al. (2014) reported a success rate of 73% (22/30) in diagnosing patients with sporadic cerebellar ataxia. Reanalysis of the undiagnosed exomes and coupling WES with other genetic approaches, such as mRNA sequencing and CNV detection, also increase WES's diagnostic yield (Deelen et al., 2019; Jalkh et al., 2019; Fung et al., 2020; Matalonga et al., 2020). CNVs are common in some SCD subtypes, e.g., SPG4 (OMIM

# 182601) and SCA15 (OMIM # 606658; Depienne et al., 2007; Hsiao et al., 2017). They can be looked for in short-reads sequencing data, but with variable rates according to the used algorithms and the library preparation, making comparative genomic hybridization (CGH) arrays or multiplex-ligation-probe-amplification (MLPA) favored to detect CNV. One study reported a rise of 7% in the diagnostic success rate upon reanalyzing the WES data of a cohort of SCD patients (Ngo et al., 2020). Recent reports advocate for a periodic reanalysis of WES data (Jalkh et al., 2019; Liu et al., 2019; Fung et al., 2020; Ngo et al., 2020). The 5-year cumulative increase in the diagnostic success rate upon WES data reanalysis ranges from 12 to 22% (Liu et al., 2019).

### Whole-Genome Sequencing

Whole-genome sequencing is the most comprehensive tool to study the genome (Gao and Smith, 2020). A significant advantage of WGS is its ability to interrogate intergenic and intronic regions that constitute most of the human genome. This ability is a double-edged sword as the interpretation of intergenic and intronic variants is hard compared to the interpretation of exonic variants (Gao and Smith, 2020). For that reason, most of the time, WGS is interrogated for exonic sequences first, and often, limited to them (Rexach et al., 2019). WGS has a lower but more uniform coverage compared to WES (Posey, 2019). This uniformity in coverage of exonic sequences has resulted in better detection of GC-rich regions, e.g., at the first exons, and CNVs, compared to WES (Posey, 2019; Rexach et al., 2019).

To date, WGS has been involved in the discovery of three SCD genes. Melo et al. (2015) identified a homozygous 216-bp deletion in the non-coding sequence upstream to the *KLC2* gene in patients with SPG, optic atrophy, and neuropathy (SPOAN). Melo et al. (2015) identified this 216-bp deletion in screening a series of 275 patients (273 from Brazil and two from Egypt) with SPOAN and validated its pathogenicity in a zebrafish model. The candidate region containing *KLC2* was refined using linkage analysis and WES (Macedo-Souza et al., 2009; Melo et al., 2015). Similarly, Cortese et al. (2019) identified bi-allelic intronic expansions in *RFC1* in 56 patients from the United Kingdom, Italy, and Brazil with the CANVAS phenotype using a WGS-based approach focused to the linked-chromosomal region. In another study, Wagner et al. (2019) using WGS, identified a biallelic splice-site variant in the *RNF170* gene in a family with early-onset HSP complicated with peripheral neuropathy. Wagner et al. (2019) corroborated their finding by identifying additional patients with pathogenic *RNF170* variants on the GENESIS platform<sup>1</sup> of shared sequencing data (Gonzalez et al., 2015).

Studies that follow the WGS-based approach to identify novel genes generally use other techniques, such as whole-genome linkage analysis, to focus the search for candidate variants. Indeed, the bottleneck of the WGS-based approach is the large number of variants that can be identified. Multiomics integration will ameliorate the interpretation of the functional

and regulatory effects of variants identified by NGS, especially WGS (Agrawal et al., 2014; Posey, 2019). Furthermore, long-read WGS is probably a promising new tool for diagnosing genetic diseases for the near future with great hope on detection of nucleotide repeats as well (Posey, 2019; Wenger et al., 2019).

## DISCUSSION AND FUTURE DIRECTIONS

### The Need for Combined Approaches

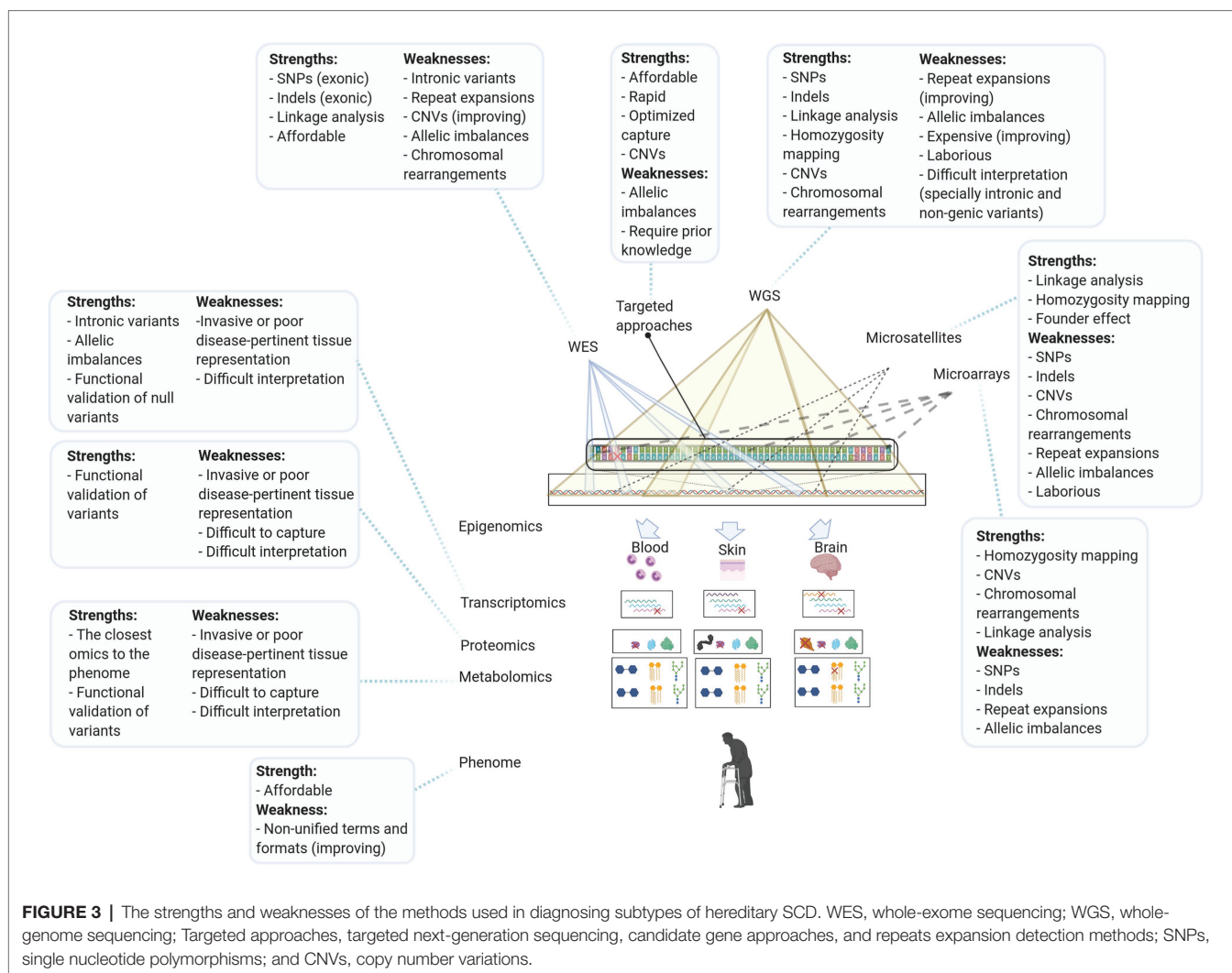
Genome-wide approaches have proved the most successful in identifying new SCD subtypes, having identified 86.4% (190/220) of the known SCD subtypes. These approaches have made use of linkage analysis, WES, WGS, and DNA microarrays. Linkage analysis with candidate gene screening dominated the discovery of SCD genes in the past, often thanks to prior knowledge of these genes' functions and expression profiles, their products, the pathways they serve, or pertinent animal models. WES is now dominating SCD gene discovery and we can expect that WGS will dominate the discovery of the genes causing Mendelian diseases in the future, probably within integrated multi-omics strategies to reduce the number of candidate variants (Posey, 2019; Kerr et al., 2020).

Indeed, approaches that utilized more than one genetic screening tool identified 55.9% (123/220) of SCD subtypes compared to the 43.2% (95/220) identified by single tools. Historically, the most successful approaches in identifying SCD genes were those involving linkage analysis and exome sequencing. However, the use of NGS to diagnose rare diseases has been accompanied by a shift away from studying single large multiplex pedigrees with a restricted phenotype toward studying phenotypically diverse multiple small pedigrees and sporadic cases as well (Wright et al., 2018; Claussnitzer et al., 2020). In our opinion, WES and SNP arrays are the best combination to accelerate the discovery of SCD genes, since SNP arrays can be used in detecting CNVs, chromosomal aberrations, and linkage/homozygosity regions (Levy and Burnside, 2019). SNP arrays are also the most widely used tools for studying complex traits by genome-wide association studies (Tam et al., 2019). WGS could achieve all these goals, in an even better way, and eliminate the need to couple WES and SNP arrays (Figure 3). However, at the present time, the use of WGS on a broad scale is prohibited by its relatively high costs and the technical difficulties in storing, analyzing, and sharing WGS data (Suwinski et al., 2019; Tam et al., 2019).

### Emerging Approaches

Several tools and developments in the genome-wide approaches have emerged to have inherent potentials for augmenting the discovery of new Mendelian diseases. These advancements include, but not limited to, the optical mapping methods for detecting genomic rearrangements (e.g., Bionano optical mapping), long-read genomic sequencing valuable for detecting copy number variants and nucleotide expansions (e.g., Oxford Nanopore or PacBio technologies), and mosaicism detection tools (e.g., MosaicForecast; Lam et al., 2012; Jain et al., 2018;

<sup>1</sup><https://www.tgp-foundation.org/genesis-log-in>



Dou et al., 2020). Structural rearrangements, CNVs, sequence expansions, and somatic mutations account for a significant proportion of the missing heritability in Mendelian diseases (Maroille and Tarailo-Graovac, 2019).

The comprehensiveness of WES and WGS should not lead us to underestimate the role of proper clinical phenotyping in diagnosing genetic diseases (Pena et al., 2018). Tools that incorporate phenotypic data in NGS analysis pipelines are under development and have been used with success (Thuriot et al., 2018; Zhao et al., 2020). These include tools that empower the clinicians to have a central role in prioritizing variants, e.g., tools that apply gene-pertinence metrics (Segal et al., 2020). Proper clinical phenotyping is of value to increase the diagnostic rate of rare diseases (Bhatia et al., 2020). The high degree of overlap between various neurodegenerative disorders, including SCD, calls for better documentation, and follow-up for patients (Klebe et al., 2015; Parodi et al., 2018). This is clearly seen by the constant identification of more genes involved in other diseases and found mutated in SCD as well. The reverse is also true and illustrates the clinical overlap with various diseases. International registries, such as SCA Global

and ARCA Global<sup>2</sup> and SPATAX,<sup>3</sup> have been launched to meet the need for better documentation of SCD cases. A prerequisite for international patient registries, besides appropriate research ethics policies, is the development of unified common disability scales to facilitate analysis and comparison (Trouillas et al., 1997; Lin et al., 2020). Many disability scales have been developed for SCD, the most commonly used scales being the scale for the assessment and rating of ataxia (SARA) and the International Cooperative Ataxia Rating Scale (ICARS; Perez-Lloret et al., 2020). Disability scales do not only measure disability; unlike what the name implies, but they can also measure the response to therapeutic interventions including attempts to preserve motor capacities of patients through alternative treatment options, e.g., dance, video game-based coordinative training (exergames), etc. (Ilg et al., 2012).

On the other hand, clinical phenotypes will, in many cases, evolve over time, and this needs to be considered when analyzing NGS data (Liu et al., 2019; Bhatia et al., 2020).

<sup>2</sup><http://ataxia-global-initiatives.net/>

<sup>3</sup><https://spatex.wordpress.com/>

Artificial intelligence tools have recently been employed in phenotyping, known as next-generation phenotyping, with good results, especially in neurological diseases (Elmas and Gogus, 2020; Pode-Shakked et al., 2020).

Experimental forward genetics has identified a few SCD genes. Forward genetics investigates the genetic basis of a particular phenotype, in contrast to reverse genetics, which investigates the phenotypic consequences of altering a particular gene (Moresco et al., 2013). The current development in forward genetics and its equipment with NGS tools could further enhance the application of forward genetics for discovering new genes implicated in Mendelian diseases (Singh, 2020). Furthermore, gene essentiality screens on model organisms and cell lines of different origins are gaining momentum and are progressively integrated into gene-prioritization steps in NGS analysis pipelines (Cacheiro et al., 2020).

## Remaining Challenges

There is still a pressing need for biomarkers to complement the advances in SCD genomics and phenomics. The role of biomarkers is not limited to diagnosis, as biomarkers can objectively measure disease evolution, prognosis, and response to therapy (Gülbakan et al., 2016). Currently, the means for validating the biological impact of candidate pathogenic SCD variants in the clinical setting are not available except for a few SCD subtypes, e.g., SPG5A caused by mutations in the *CYP7B1* gene (27-OH cholesterol), SPG9A (OMIM # 601162), and SPG9B (OMIM # 616586) caused by mutations in the *ALDH18A1* gene (amino-acid levels in blood), and ataxia with isolated vitamin E deficiency (OMIM # 277460) caused by mutations in the *TTPA* gene (Ouahchi et al., 1995; Coutelier et al., 2015; Marelli et al., 2018).

The rapidly evolving field of metabolomics is expected to contribute to the SCD biomarkers list (Wishart, 2016; Yang et al., 2019). Metabolomics is analyzing the body metabolites in a comprehensive, yet systematic approach (Gülbakan et al., 2016). Metabolic disturbances are core pathological mechanisms in many SCD subtypes (Blackstone, 2018; Darios et al., 2020). Consequently, metabolomics will not only provide biomarkers but also will shed light on the SCD biochemistry and provide an objective means for classifying SCD subtypes (Wishart, 2016; del Mar Amador et al., 2018; Yang et al., 2019). It can also unravel the overlapping phenotypes caused by mutations in a single gene, a pattern common in neurological diseases, including SCD (Esterhuizen et al., 2021). Developing an objective SCD nosology is a prerequisite for advancing SCD research and therapeutics (Elsayed et al., 2019).

Metabolomics is the closest omics layer to the phenotype and could have central roles in validating the newly discovered Mendelian genes (Feussner and Polle, 2015; Graham et al., 2018; Almontashiri et al., 2020). Metabolomics studies are conducted using two main approaches: a hypothesis-directed approach that studies a predetermined set of metabolites (targeted metabolomics) and a hypothesis-free comprehensive approach that aims to analyze the whole set of metabolites within a system (untargeted or global metabolomics; Almontashiri et al., 2020; Chen et al., 2020). Global metabolomics is gaining momentum over targeted approaches in clinical settings

(Almontashiri et al., 2020). However, neither approach is suitable yet for the *ab initio* discovery of new genes; rather, global metabolomics can be integrated with exome and genome sequencing to enhance new genes discovery (Graham et al., 2018; Almontashiri et al., 2020).

Multiple hurdles are facing the application of metabolomics in Mendelian genes discovery. Firstly, the human metabolome is made of thousands of compounds influenced by the genome but also by a plethora of non-genomic determinants, including the microbiome, lifestyle, environment, and others (Bar et al., 2020). Secondly, a proportion of the human metabolome is not identified yet and there is no single tool currently capable of capturing the whole metabolome (Graham et al., 2018). Furthermore, there is no congruence between the concentration of certain metabolites in the blood and the brain (Yang et al., 2019). Lastly, the limited number of patients with rare diseases, particularly those in whom new phenotypes are discovered, complicates the analysis of metabolomics data and the differentiation between genuine signals and background noise (Graham et al., 2018; Yang et al., 2019). However, recently, some promising solutions have emerged to overcome those hurdles.

Multiple repositories were launched to enhance metabolomics data sharing, and, in parallel, tools that integrate the data from different repositories were recently developed (Krassowski et al., 2020; Palermo et al., 2020). Examples for such repositories include MetaboLights, GNPS-MASSIVE, and Metabolomics Workbench (Haug et al., 2013; Sud et al., 2016; Wang et al., 2016). Animal models can be used to alleviate the need for human tissues to study the metabolome (Kilk, 2019). Even more promising are the human organoids, which are more representative substitutes for the human brain tissues (Kim et al., 2020). To a large extent, organoids can also solve the scarcity of patients with rare diseases (Kim et al., 2020).

Other omics layers, particularly the transcriptome and the proteome, can be exploited in the quest for identifying new SCD genes with the limitation of the access to the brain tissue in most cases. A substantial proportion of the undiagnosed Mendelian diseases are caused by variants impacting RNA expression and/or regulation, including deep intronic variants (Gonorazky et al., 2019). Transcriptome sequencing (RNA-seq) is deemed to enhance identifying Mendelian variants by ~10–30% (Kremer et al., 2017; Gonorazky et al., 2019). Allelic imbalances in transcriptomics data can identify imprinting, uni-parental disomy, and X chromosome inactivation (Gonorazky et al., 2019; Yépez et al., 2021). However, the tissue transcriptome does not necessarily reflect the proteome, metabolome, and phenome (Wang et al., 2019). In the abnormal tissues, the discrepancy between the transcriptome and the proteome can be more evident depending on the underlying disease's pathophysiology and the type of the disease-causing variants (Roos et al., 2018). Although the proteome is closer to the phenome than the transcriptome, it is more easier to recover data by RNA-seq than proteome-capturing approaches due to the higher depth of RNA-seq and the much wider proteins' dynamic range of expression (Feussner and Polle, 2015; Wang et al., 2019). Moreover, ~13% of the human proteome is still missing and not captured (Omenn et al., 2017; Wang et al., 2019).



A holistic approach that integrates data obtained from multiple omics layers (multi-omics approach) has a higher biological reliability and can substantially enhance new Mendelian genes' discovery than the single-omics approaches (Labory et al., 2020). The multi-omics have been successfully used in many diseases, including neurological diseases, and is thus a promising approach to enhance the discovery of new SCD genes (Crowther et al., 2018; Kerr et al., 2020; Labory et al., 2020).

In the current big data era, tools that integrate the data generated at different biological scales, namely multi-omic tools, are rapidly emerging (Anderson et al., 2019; Kerr et al., 2020; Subramanian et al., 2020). This promotes collaborative studies. In some instances, establishing associations between variants, especially non-exonic variants, and diseases, required a large number of patients and controls. Involving patients from different genetic backgrounds has aided the refinement of the loci associated with some SCD genes. An increasing number of consortia and collaborative platforms have been launched to enhance the discovery of new genes by enabling the sharing of sequence data, e.g., GENESIS platform; variants and phenotypes, e.g., GeneMatcher; and connecting geneticists and biologists to enhance validating new candidate genes, e.g., the Canadian Rare Diseases Models and Mechanisms Network (Gonzalez et al., 2015; Sobreira et al., 2015; Boycott et al., 2020). Larger-scale collaborations are essential to explore a very challenging aspect of all neurogenetic conditions, the clinical variability. Modifier genes can explain part of the phenotypic variability commonly seen between patients, as shown in autosomal dominant SCAs (Du Montcel et al., 2014).

Sporadic cases remain a challenge in these diseases. They are often the most frequent in clinical practice and their analysis possibility is limited. Collaborations, trio analysis, and multi-omic approaches can assist in solving them. In our experience, 10% of sporadic cases can be explained by *de novo* variants or inherited variants due to incomplete penetrance, adoption, family censure, etc. We expect multigenic alterations or somatic variants to explain as well part of the missing heritability in some of the patients with sporadic forms.

Finally, one of the bottlenecks in diagnosing SCD is the functional validation of the variants of unknown significance (VUS). Three-dimensional modeling and functional enzymatic assays can be precious, as shown for variants in the *CYP2U1* gene, the gene associated with SPG56 (OMIM # 615030; Durand et al., 2018). Small animal models such as *C. elegans* and zebrafish can also be used (Martin et al., 2013; Novarino et al., 2014; Chiba et al., 2019). Moreover, organoids emerged as promising models to establish pathogenicity and circumvent the difficulty in obtaining brain tissues (Kim et al., 2020). Without such systematic studies, we can reasonably think

that a significant part of variants are not really causative which may limit the finding of phenotype-genotype correlations. In that respect, the report of second mutated families is required for multiple clinico genetic entities, including SPG59, 60, 65, 67, 68, 69, 70, and 71, found in single families yet (Novarino et al., 2014).

## CONCLUSION

In conclusion, the toolbox for diagnosing Mendelian diseases is continuously expanding, including multiple "out of the box" approaches. These valuable utensils and approaches (Figure 3) have solved and deemed to continue solving some of the Mendelian diseases diagnostic odysseys. This review highlighted some of the approaches successfully used to discover SCD genes and shed light on some of the promising resources and approaches expected to play central roles in diagnosing SCD soon. Discovering novel genes associated with SCD is moving toward integrated multi-omic research using innovative technologies. We hypothesize that many neurogenetic entities will join the spectrum of SCD, necessitating newer and more comprehensive nosologies at the SCD, and likely at the neurogenetic level as well.

## AUTHOR CONTRIBUTIONS

AY and GS designed, wrote, and critically revised this review article. All authors contributed to the article and approved the submitted version.

## FUNDING

GS is financially supported by a European Union grant (grant 779257 Solve-RD from the Horizon 2020 Research and Innovation Program) and by Grants from the Spastic Paraplegia Foundation (US), the HSP-ASL-France and the association Connaitre les Syndromes Cérébelleux. AY is supported by a grant from the Ministry of Higher Education in Sudan and the French embassy in Sudan.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.638730/full#supplementary-material>

## REFERENCES

- Abou Jamra, R., Philippe, O., Raas-Rothschild, A., Eck, S. H., Graf, E., Buchert, R., et al. (2011). Adaptor protein complex 4 deficiency causes severe autosomal-recessive intellectual disability, progressive spastic paraplegia, shy character, and short stature. *Am. J. Hum. Genet.* 88, 788–795. doi: 10.1016/j.ajhg.2011.04.019
- Agrawal, P. B., Joshi, M., Marinakis, N. S., Schmitz-Abe, K., Ciarlini, P. D. S. C., Sargent, J. C., et al. (2014). Expanding the phenotype associated with the NEFL mutation neuromuscular disease in a family with overlapping myopathic and neurogenic findings. *JAMA Neurol.* 71, 1413–1420. doi: 10.1001/jamaneurol.2014.1432
- Almontashiri, N. A. M., Zha, L., Young, K., Law, T., Kellogg, M. D., Bodamer, O. A., et al. (2020). Clinical validation of targeted and untargeted metabolomics

- testing for genetic disorders: a 3 year comparative study. *Sci. Rep.* 10, 1–8. doi: 10.1038/s41598-020-66401-2
- Anderson, D., Baynam, G., Blackwell, J. M., and Lassmann, T. (2019). Personalised analytics for rare disease diagnostics. *Nat. Commun.* 10, 1–8. doi: 10.1038/s41467-019-13345-5
- Bahlo, M., Bennett, M. F., Degorski, P., Tankard, R. M., Delatycki, M. B., and Lockhart, P. J. (2018). Recent advances in the detection of repeat expansions with short-read next-generation sequencing. *F1000Research* 7:736. doi: 10.12688/f1000research.13980.1
- Bar, N., Korem, T., Weissbrod, O., Zeevi, D., Rothschild, D., Leviatan, S., et al. (2020). A reference map of potential determinants for the human serum metabolome. *Nature* 588, 135–140. doi: 10.1038/s41586-020-2896-2
- Bell, S. M., Schreiner, C. M., Schultheis, P. J., Miller, M. L., Evans, R. L., Vorhees, C. V., et al. (1999). Targeted disruption of the murine *Nhe1* locus induces ataxia, growth retardation, and seizures. *Am. J. Phys.* 276, C788–C795. doi: 10.1152/ajpcell.1999.276.4.c788
- Bhatia, N. S., Lim, J. Y., Bonnard, C., Kuan, J. L., Brett, M., Wei, H., et al. (2020). Singapore undiagnosed disease program: genomic analysis aids diagnosis and clinical management. *Arch. Dis. Child.* 106, 31–37. doi: 10.1136/archdischild-2020-319180
- Blackstone, C. (2018). Converging cellular themes for the hereditary spastic paraplegias. *Curr. Opin. Neurobiol.* 51, 139–146. doi: 10.1016/j.conb.2018.04.025
- Boutry, M., Morais, S., and Stevanin, G. (2019). Update on the genetics of spastic paraplegias. *Curr. Neurol. Neurosci. Rep.* 19:18. doi: 10.1007/s11910-019-0930-2
- Boycott, K. M., Campeau, P. M., Howley, H. E., Pavlidis, P., Rogic, S., Oriol, C., et al. (2020). The Canadian rare diseases models and mechanisms (RDMM) network: connecting understudied genes to model organisms. *Am. J. Hum. Genet.* 106, 143–152. doi: 10.1016/j.ajhg.2020.01.009
- Boycott, K. M., Rath, A., Chong, J. X., Hartley, T., Alkuraya, F. S., Baynam, G., et al. (2017). International cooperation to enable the diagnosis of all rare genetic diseases. *Am. J. Hum. Genet.* 100, 695–705. doi: 10.1016/j.ajhg.2017.04.003
- Bruse, E., Maat-Kievit, J. A., and van Swieten, J. C. (2007). Diagnosis and management of early- and late-onset cerebellar ataxia. *Clin. Genet.* 71, 12–24. doi: 10.1111/j.1399-0004.2006.00722.x
- Burgess, D. L., Jones, J. M., Meisler, M. H., and Noebels, J. L. (1997). Mutation of the  $Ca^{2+}$  channel  $\beta$  subunit gene *Cchb4* is associated with ataxia and seizures in the lethargic (lh) mouse. *Cell* 88, 385–392. doi: 10.1016/S0092-8674(00)81877-2
- Cacheiro, P., Muñoz-Fuentes, V., Murray, S. A., Dickinson, M. E., Bucan, M., Nutter, L. M. J., et al. (2020). Human and mouse essentiality screens as a resource for disease gene discovery. *Nat. Commun.* 11:655. doi: 10.1038/s41467-020-14284-2
- Cagnoli, C., Brussino, A., Mancini, C., Ferrone, M., Orsi, L., Salmin, P., et al. (2018). Spinocerebellar ataxia tethering PCR: a rapid genetic test for the diagnosis of spinocerebellar ataxia types 1, 2, 3, 6, and 7 by PCR and capillary electrophoresis. *J. Mol. Diagn.* 20, 289–297. doi: 10.1016/j.jmoldx.2017.12.006
- Campuzano, V., Montermini, L., Moltò, M. D., Pianese, L., Cossée, M., Cavalcanti, F., et al. (1996). Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* 271, 1423–1427. doi: 10.1126/science.271.5254.1423
- Caspar, S. M., Dubacher, N., Kopps, A. M., Meienberg, J., Henggeler, C., and Matyas, G. (2018). Clinical sequencing: from raw data to diagnosis with lifetime value. *Clin. Genet.* 93, 508–519. doi: 10.1111/cge.13190
- Chen, L., Zhong, F., and Zhu, J. (2020). Bridging targeted and untargeted mass spectrometry-based metabolomics via hybrid approaches. *Meta* 10, 1–19. doi: 10.3390/metabo10090348
- Chiba, K., Takahashi, H., Chen, M., Obinata, H., Arai, S., Hashimoto, K., et al. (2019). Disease-associated mutations hyperactivate KIF1A motility and anterograde axonal transport of synaptic vesicle precursors. *Proc. Natl. Acad. Sci. U. S. A.* 116, 18429–18434. doi: 10.1073/pnas.1905690116
- Ciaccio, C., Pantaleoni, C., Bulgheroni, S., Sciacca, F., and D'Arrigo, S. (2020). Chromosomal microarray analysis has a poor diagnostic yield in children with developmental delay/intellectual disability when concurrent cerebellar anomalies are present. *Cerebellum* 19, 629–635. doi: 10.1007/s12311-020-01145-3
- Clark, M. M., Stark, Z., Farnaes, L., Tan, T. Y., White, S. M., Dimmock, D., et al. (2018). Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. *NPJ Genomic Med.* 3:16. doi: 10.1038/s41525-018-0053-8
- Claussnitzer, M., Cho, J. H., Collins, R., Cox, N. J., Dermitzakis, E. T., Hurles, M. E., et al. (2020). A brief history of human disease genetics. *Nature* 577, 179–189. doi: 10.1038/s41586-019-1879-7
- Cortese, A., Simone, R., Sullivan, R., Vandrovцова, J., Tariq, H., Yan, Y. W., et al. (2019). Biallelic expansion of an intronic repeat in *RFC1* is a common cause of late-onset ataxia. *Nat. Genet.* 51, 649–658. doi: 10.1038/s41588-019-0372-4
- Coutelier, M., Goizet, C., Durr, A., Habarou, F., Morais, S., Dionne-Laporte, A., et al. (2015). Alteration of ornithine metabolism leads to dominant and recessive hereditary spastic paraplegia. *Brain* 138, 2191–2205. doi: 10.1093/brain/awv143
- Coutelier, M., Hammer, M. B., Stevanin, G., Monin, M. L., Davoine, C. S., Mochel, F., et al. (2018). Efficacy of exome-targeted capture sequencing to detect mutations in known cerebellar ataxia genes. *JAMA Neurol.* 75, 591–599. doi: 10.1001/jamaneurol.2017.5121
- Crowther, L. M., Poms, M., and Pecklo, B. (2018). Multiomics tools for the diagnosis and treatment of rare neurological disease. *J. Inherit. Metab. Dis.* 41, 425–434. doi: 10.1007/s10545-018-0154-7
- Darios, F., Mochel, F., and Stevanin, G. (2020). Lipids in the physiopathology of hereditary spastic paraplegias. *Front. Neurosci.* 14:74. doi: 10.3389/fnins.2020.00074
- David, G., Abbas, N., Stevanin, G., Dürr, A., Yvert, G., Cancel, G., et al. (1997). Cloning of the *SCA7* gene reveals a highly unstable CAG repeat expansion. *Nat. Genet.* 17, 65–70. doi: 10.1038/ng0997-65
- Deelen, P., van Dam, S., Herkert, J. C., Karjalainen, J. M., Brugge, H., Abbott, K. M., et al. (2019). Improving the diagnostic yield of exome-sequencing by predicting gene-phenotype associations using large-scale gene expression analysis. *Nat. Commun.* 10, 1–13. doi: 10.1038/s41467-019-10649-4
- del Mar Amador, M., Colsch, B., Lamari, F., Jardel, C., Ichou, F., Rastetter, A., et al. (2018). Targeted versus untargeted omics—the CAFSA story. *J. Inherit. Metab. Dis.* 41, 447–456. doi: 10.1007/s10545-017-0134-3
- Depienne, C., Fedirko, E., Forlani, S., Cazeneuve, C., Ribai, P., Feki, I., et al. (2007). Exon deletions of *SPG4* are a frequent cause of hereditary spastic paraplegia. *J. Med. Genet.* 44, 281–284. doi: 10.1136/jmg.2006.046425
- De Roeck, A., De Coster, W., Bossaerts, L., Cacace, R., De Pooter, T., Van Dongen, J., et al. (2019). NanoSatellite: accurate characterization of expanded tandem repeat length and sequence through whole genome long-read sequencing on PromethION. *Genome Biol.* 20:239. doi: 10.1186/s13059-019-1856-3
- De Silva, R., Greenfield, J., Cook, A., Bonney, H., Vallortigara, J., Hunt, B., et al. (2019). Guidelines on the diagnosis and management of the progressive ataxias. *Orphanet J. Rare Dis.* 14:51. doi: 10.1186/s13023-019-1013-9
- Devanna, P., Chen, X. S., Ho, J., Gajewski, D., Smith, S. D., Gialluisi, A., et al. (2018). Next-gen sequencing identifies non-coding variation disrupting miRNA-binding sites in neurological disorders. *Mol. Psychiatry* 23, 1375–1384. doi: 10.1038/mp.2017.30
- Dou, Y., Kwon, M., Rodin, R. E., Cortés-Ciriano, I., Doan, R., Luquette, L. J., et al. (2020). Accurate detection of mosaic variants in sequencing data without matched controls. *Nat. Biotechnol.* 38, 314–319. doi: 10.1038/s41587-019-0368-8
- Duchesne, A., Vaiman, A., Frah, M., Floriot, S., Legoueix-Rodriguez, S., Desmazières, A., et al. (2018). Progressive ataxia of Charolais cattle highlights a role of *KIF1C* in sustainable myelination. *PLoS Genet.* 14:e1007550. doi: 10.1371/journal.pgen.1007550
- Du Montcel, S. T., Durr, A., Bauer, P., Figueroa, K. P., Ichikawa, Y., Brussino, A., et al. (2014). Modulation of the age at onset in spinocerebellar ataxia by CAG tracts in various genes. *Brain* 137, 2444–2455. doi: 10.1093/brain/awu174
- Durand, C. M., Dhers, L., Tesson, C., Tessa, A., Fouillen, L., Jacquéré, S., et al. (2018). *CYP2U1* activity is altered by missense mutations in hereditary spastic paraplegia 56. *Hum. Mutat.* 39, 140–151. doi: 10.1002/humu.23359
- Elmas, M., and Gogus, B. (2020). Success of face analysis technology in rare genetic diseases diagnosed by whole-exome sequencing: a single-center experience. *Mol. Syndromol.* 11, 4–14. doi: 10.1159/000505800
- Elsayed, L. E. O., Eltazi, I. Z. M., Ahmed, A. E. M., and Stevanin, G. (2019). Hereditary spastic paraplegias: time for an objective case definition and a

- new nosology for neurogenetic disorders to facilitate biomarker/therapeutic studies. *Expert. Rev. Neurother.* 19, 409–415. doi: 10.1080/14737175.2019.1608824
- Elsayed, L. E. O., Mohammed, I. N., Hamed, A. A. A., Elseed, M. A., Johnson, A., Mairey, M., et al. (2016). Hereditary spastic paraplegias: identification of a novel SPG57 variant affecting TFG oligomerization and description of HSP subtypes in Sudan. *Eur. J. Hum. Genet.* 25, 100–110. doi: 10.1038/ejhg.2016.108
- Engert, J. C., Bérubé, P., Mercier, J., Doré, C., Lepage, P., Ge, B., et al. (2000). ARSACS, a spastic ataxia common in northeastern Québec, is caused by mutations in a new gene encoding an 11.5-kb ORF. *Nat. Genet.* 24, 120–125. doi: 10.1038/72769
- Erlich, Y., Edvardson, S., Hodges, E., Zenvirt, S., Thekkat, P., Shaag, A., et al. (2011). Exome sequencing and disease-network analysis of a single family implicate a mutation in KIF1A in hereditary spastic paraparesis. *Genome Res.* 21, 658–664. doi: 10.1101/gr.117143.110
- Escayg, A., De Waard, M., Lee, D. D., Bichet, D., Wolf, P., Mayer, T., et al. (2000). Coding and noncoding variation of the human calcium-channel  $\beta$ 4-subunit gene CACNB4 patients with idiopathic generalized epilepsy and episodic ataxia. *Am. J. Hum. Genet.* 66, 1531–1539. doi: 10.1086/302909
- Esterhuizen, K., Lindeque, J. Z., Mason, S., van der Westhuizen, F. H., Rodenburg, R. J., de Laat, P., et al. (2021). One mutation, three phenotypes: novel metabolic insights on MELAS, MIDD and myopathy caused by the m.3243A > G mutation. *Metabolism* 17, 1–16. doi: 10.1007/s11306-020-01769-w
- Feussner, I., and Polle, A. (2015). What the transcriptome does not tell—proteomics and metabolomics are closer to the plants' patho-phenotype. *Curr. Opin. Plant Biol.* 26, 26–31. doi: 10.1016/j.pbi.2015.05.023
- Fink, J. K. (1997). Advances in hereditary spastic paraplegia. *Curr. Opin. Neurol.* 10, 313–318. doi: 10.1097/00019052-199708000-00006
- Fogel, B. L., Lee, H., Deignan, J. L., Strom, S. P., Kantarci, S., Wang, X., et al. (2014). Exome sequencing in the clinical diagnosis of sporadic or familial cerebellar ataxia. *JAMA Neurol.* 71, 1237–1246. doi: 10.1001/jamaneurol.2014.1944
- Friesema, E. C. H., Grueters, P. A., Biebermann, H., Krude, H., von Moers, A., Reeser, M., et al. (2004). Association between mutations in a thyroid hormone transporter and severe X-linked psychomotor retardation. *Lancet* 364, 1435–1437. doi: 10.1016/S0140-6736(04)17226-7
- Fung, J. L. F., Yu, M. H. C., Huang, S., Chung, C. C. Y., Chan, M. C. Y., Pajusalu, S., et al. (2020). A three-year follow-up study evaluating clinical utility of exome sequencing and diagnostic potential of reanalysis. *NPJ Genomic Med.* 5:37. doi: 10.1038/s41525-020-00144-x
- Galatolo, D., Tessa, A., Filla, A., and Santorelli, F. M. (2018). Clinical application of next generation sequencing in hereditary spinocerebellar ataxia: increasing the diagnostic yield and broadening the ataxia-spasticity spectrum. A retrospective analysis. *Neurogenetics* 19, 1–18. doi: 10.1007/s10048-017-0532-6
- Gao, G., and Smith, D. I. (2020). Clinical massively parallel sequencing. *Clin. Chem.* 66, 77–88. doi: 10.1373/clinchem.2019.303305
- Gonorazky, H. D., Naumenko, S., Ramani, A. K., Nelakuditi, V., Mashouri, P., Wang, P., et al. (2019). Expanding the boundaries of RNA sequencing as a diagnostic tool for rare mendelian disease. *Am. J. Hum. Genet.* 104, 466–483. doi: 10.1016/j.ajhg.2019.01.012
- Gonzalez, M., Falk, M. J., Gai, X., Postrel, R., Schüle, R., and Zuchner, S. (2015). Innovative genomic collaboration using the genesis (GEM.App) platform. *Hum. Mutat.* 36, 950–956. doi: 10.1002/humu.22836
- Graham, E., Lee, J., Price, M., Tarailo-Graovac, M., Matthews, A., Engelke, U., et al. (2018). Integration of genomics and metabolomics for prioritization of rare disease variants: a 2018 literature review. *J. Inherit. Metab. Dis.* 41, 435–445. doi: 10.1007/s10545-018-0139-6
- Guerguelcheva, V., Azmanov, D. N., Angelicheva, D., Smith, K. R., Chamova, T., Florez, L., et al. (2012). Autosomal-recessive congenital cerebellar ataxia is caused by mutations in metabotropic glutamate receptor 1. *Am. J. Hum. Genet.* 91, 553–564. doi: 10.1016/j.ajhg.2012.07.019
- Guissart, C., Li, X., Leheup, B., Drouot, N., Montaut-Verier, B., Raffo, E., et al. (2015). Mutation of SLC9A1, encoding the major Na<sup>+</sup>/H<sup>+</sup> exchanger, causes ataxia-deafness Lichtenstein-Knorr syndrome. *Hum. Mol. Genet.* 24, 463–470. doi: 10.1093/hmg/ddu461
- Gülbakan, B., Özgül, R. K., Yüzbaşıoğlu, A., Kohl, M., Deigner, H. P., and Özgüç, M. (2016). Discovery of biomarkers in rare diseases: innovative approaches by predictive and personalized medicine. *EPMA J.* 7:24. doi: 10.1186/s13167-016-0074-2
- Hagerman, R. J., Leehey, M., Heinrichs, W., Tassone, F., Wilson, R., Hills, J., et al. (2001). Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. *Neurology* 57, 127–130. doi: 10.1212/WNL.57.1.127
- Hansen, M. C., Haferlach, T., and Nyvold, C. G. (2020). A decade with whole exome sequencing in haematology. *Br. J. Haematol.* 188, 367–382. doi: 10.1111/bjh.16249
- Harding, A. E. (1993). Hereditary spastic paraplegias. *Semin. Neurol.* 13, 333–336. doi: 10.1055/s-2008-1041143
- Haug, K., Salek, R. M., Conesa, P., Hastings, J., De Matos, P., Rijnbeek, M., et al. (2013). MetaboLights: an open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucleic Acids Res.* 41, D781–D786. doi: 10.1093/nar/gks1004
- Hershenson, J., Mencacci, N. E., Davis, M., MacDonald, N., Trabzuni, D., Ryten, M., et al. (2013). Mutations in the autoregulatory domain of  $\beta$ -tubulin 4a cause hereditary dystonia. *Ann. Neurol.* 73, 546–553. doi: 10.1002/ana.23832
- Holt, I. J., Harding, A. E., Petty, R. K. H., and Morgan-Hughes, J. A. (1990). A new mitochondrial disease associated with mitochondrial DNA heteroplasmy. *Am. J. Hum. Genet.* 46, 428–433.
- Hsiao, C. T., Liu, Y. T., Liao, Y. C., Hsu, T. Y., Lee, Y. C., and Soong, B. W. (2017). Mutational analysis of ITPR1 in a Taiwanese cohort with cerebellar ataxias. *PLoS One* 12:e0187503. doi: 10.1371/journal.pone.0187503
- Ikeda, Y., Dick, K. A., Weatherspoon, M. R., Gincel, D., Armbrust, K. R., Dalton, J. C., et al. (2006). Spectrin mutations cause spinocerebellar ataxia type 5. *Nat. Genet.* 38, 184–190. doi: 10.1038/ng1728
- Ilg, W., Schatton, C., Schicks, J., Giese, M. A., Schöls, L., and Synofzik, M. (2012). Video game-based coordinative training improves ataxia in children with degenerative ataxia. *Neurology* 79, 2056–2060. doi: 10.1212/WNL.0b013e3182749e67
- Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., et al. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* 36, 338–345. doi: 10.1038/nbt.4060
- Jalkh, N., Corbani, S., Haidar, Z., Hamdan, N., Farah, E., Abou Ghoch, J., et al. (2019). The added value of WES reanalysis in the field of genetic diagnosis: lessons learned from 200 exomes in the Lebanese population. *BMC Med. Genet.* 12:11. doi: 10.1186/s12920-019-0474-y
- Jang, W., Kim, Y., Han, E., Park, J., Chae, H., Kwon, A., et al. (2019). Chromosomal microarray analysis as a first-tier clinical diagnostic test in patients with developmental delay/intellectual disability, autism spectrum disorders, and multiple congenital anomalies: a prospective multicenter study in Korea. *Ann. Lab. Med.* 39, 299–310. doi: 10.3343/alm.2019.39.3.299
- Jen, J. C., Wan, J., Palos, T. P., Howard, B. D., and Baloh, R. W. (2005). Mutation in the glutamate transporter EAAT1 causes episodic ataxia, hemiplegia, and seizures. *Neurology* 65, 529–534. doi: 10.1212/01.WNL.0000172638.58172.5a
- Kerr, K., McAneney, H., Smyth, L. J., Bailie, C., McKee, S., and McKnight, A. J. (2020). A scoping review and proposed workflow for multi-omic rare disease research. *Orphanet J. Rare Dis.* 15:107. doi: 10.1186/s13023-020-01376-x
- Kilk, K. (2019). Metabolomics for animal models of rare human diseases: an expert review and lessons learned. *Omi. A J. Integr. Biol.* 23, 300–307. doi: 10.1089/omi.2019.0065
- Kim, J., Koo, B. K., and Knoblich, J. A. (2020). Human organoids: model systems for human biology and medicine. *Nat. Rev. Mol. Cell Biol.* 21, 571–584. doi: 10.1038/s41580-020-0259-3
- Klebe, S., Stevanin, G., and Depienne, C. (2015). Clinical and genetic heterogeneity in hereditary spastic paraplegias: from SPG1 to SPG72 and still counting. *Rev. Neurol.* 171, 505–530. doi: 10.1016/j.neurol.2015.02.017
- Koide, R., Ikeuchi, T., Onodera, O., Tanaka, H., Igarashi, S., Endo, K., et al. (1994). Unstable expansion of CAG repeat in hereditary dentatorubral-pallidolusian atrophy (DRPLA). *Nat. Genet.* 6, 9–13. doi: 10.1038/ng0194-9
- Koob, M. D., Moseley, M. L., Schut, L. J., Benzow, K. A., Bird, T. D., Day, J. W., et al. (1999). An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). *Nat. Genet.* 21, 379–384. doi: 10.1038/7710
- Krassowski, M., Das, V., Sahu, S. K., and Misra, B. B. (2020). State of the field in multi-omics research: from computational needs to data mining and sharing. *Front. Genet.* 11:610798. doi: 10.3389/fgene.2020.610798
- Kremer, L. S., Bader, D. M., Mertes, C., Kopajtic, R., Pichler, G., Juso, A., et al. (2017). Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat. Commun.* 8:15824. doi: 10.1038/ncomms15824
- Labory, J., Fierville, M., Ait-El-Mkadem, S., Bannwarth, S., Paquis-Fluckinger, V., and Bottini, S. (2020). Multi-omics approaches to improve mitochondrial

- disease diagnosis: challenges, advances, and perspectives. *Front. Mol. Biosci.* 7:327. doi: 10.3389/fmolb.2020.590842
- Lam, E. T., Hastie, A., Lin, C., Ehrlich, D., Das, S. K., Austin, M. D., et al. (2012). Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. *Nat. Biotechnol.* 30, 771–776. doi: 10.1038/nbt.2303
- Lan, M. Y., Chang, Y. Y., Yeh, T. H., Lin, T. K., and Lu, C. S. (2017). Leukoencephalopathy with brainstem and spinal cord involvement and lactate elevation (LBSL) with a novel DARS2 mutation and isolated progressive spastic paraparesis. *J. Neurol. Sci.* 372, 229–231. doi: 10.1016/j.jns.2016.11.058
- Lee, C. E., Singleton, K. S., Wallin, M., and Faundez, V. (2020). Rare genetic diseases: nature's experiments on human development. *iScience* 23:101123. doi: 10.1016/j.isci.2020.101123
- Levy, B., and Burnside, R. D. (2019). Are all chromosome microarrays the same? What clinicians need to know. *Prenat. Diagn.* 39, 157–164. doi: 10.1002/pd.5422
- Li, S. H., McInnis, M. G., Margolis, R. L., Antonarakis, S. E., and Ross, C. A. (1993). Novel triplet repeat containing genes in human brain: cloning, expression, and length polymorphisms. *Genomics* 16, 572–579. doi: 10.1006/geno.1993.1232
- Liao, Y., Anttonen, A. K., Liukkonen, E., Gaily, E., Maljevic, S., Schubert, S., et al. (2010). SCN2A mutation associated with neonatal epilepsy, late-onset episodic ataxia, myoclonus, and pain. *Neurology* 75, 1454–1458. doi: 10.1212/WNL.0b013e3181f8812e
- Lim, J., Hao, T., Shaw, C., Patel, A. J., Szabó, G., Rual, J. F., et al. (2006). A protein-protein interaction network for human inherited ataxias and disorders of purkinje cell degeneration. *Cell* 125, 801–814. doi: 10.1016/j.cell.2006.03.032
- Lin, C.-C., Ashizawa, T., and Kuo, S.-H. (2020). Collaborative efforts for spinocerebellar ataxia research in the United States: CRC-SCA and READISCA. *Front. Neurol.* 11:902. doi: 10.3389/fneur.2020.00902
- Lipner, E. M., and Greenberg, D. A. (2018). “The rise and fall and rise of linkage analysis as a technique for finding and characterizing inherited influences on disease expression” in *Disease gene identification*. ed. J. DiStefano (Humana Press Inc.), 381–397.
- Lise, S., Clarkson, Y., Perkins, E., Kwasniewska, A., Sadighi Akha, E., Parolin Schnekenberg, R., et al. (2012). Recessive mutations in SPTBN2 implicate  $\beta$ -III spectrin in both cognitive and motor development. *PLoS Genet.* 8:e1003074. doi: 10.1371/journal.pgen.1003074
- Liu, P., Meng, L., Normand, E. A., Xia, F., Song, X., Ghazi, A., et al. (2019). Reanalysis of clinical exome sequencing data. *N. Engl. J. Med.* 380, 2478–2480. doi: 10.1056/NEJMc1812033
- Liu, Y., Zaun, H. C., Orłowski, J., and Ackerman, S. L. (2013). CHP1-mediated NHE1 biosynthetic maturation is required for purkinje cell axon homeostasis. *J. Neurosci.* 33, 12656–12669. doi: 10.1523/JNEUROSCI.0406-13.2013
- Macedo-Souza, L. I., Kok, F., Santos, S., Licinio, L., Lezirovitz, K., Cavaçana, N., et al. (2009). Spastic paraplegia, optic atrophy, and neuropathy: new observations, locus refinement, and exclusion of candidate genes. *Ann. Hum. Genet.* 73, 382–387. doi: 10.1111/j.1469-1809.2009.00507.x
- Manto, M., Gandini, J., Feil, K., and Strupp, M. (2020). Cerebellar ataxias: an update. *Curr. Opin. Neurol.* 33, 150–160. doi: 10.1097/WCO.0000000000000774
- Marelli, C., Lamari, F., Rainteau, D., Lafourcade, A., Banneau, G., Humbert, L., et al. (2018). Plasma oxysterols: biomarkers for diagnosis and treatment in spastic paraplegia type 5. *Brain* 141, 72–84. doi: 10.1093/brain/awx297
- Maroille, T., and Tarailo-Graovac, M. (2019). Uncovering missing heritability in rare diseases. *Gene* 10:275. doi: 10.3390/genes10040275
- Martin, E., Schüle, R., Smets, K., Rastetter, A., Boukhris, A., Loureiro, J. L., et al. (2013). Loss of function of glucocerebrosidase GBA2 is responsible for motor neuron defects in hereditary spastic paraplegia. *Am. J. Hum. Genet.* 92, 238–244. doi: 10.1016/j.ajhg.2012.11.021
- Martins, S., Calafell, F., Gaspar, C., Wong, V. C. N., Silveira, I., Nicholson, G. A., et al. (2007). Asian origin for the worldwide-spread mutational event in Machado-Joseph disease. *Arch. Neurol.* 64, 1502–1508. doi: 10.1001/archneur.64.10.1502
- Matalonga, L., Laurie, S., Papakonstantinou, A., Piscia, D., Mereu, E., Bullich, G., et al. (2020). Improved diagnosis of rare disease patients through systematic detection of runs of homozygosity. *J. Mol. Diagn.* 22, 1205–1215. doi: 10.1016/j.jmoldx.2020.06.008
- Mazzarotto, F., Olivotto, I., and Walsh, R. (2020). Advantages and perils of clinical whole-exome and whole-genome sequencing in cardiomyopathy. *Cardiovasc. Drugs Ther.* 34, 241–253. doi: 10.1007/s10557-020-06948-4
- Melo, U. S., Macedo-Souza, L. I., Figueiredo, T., Muotri, A. R., Gleeson, J. G., Coux, G., et al. (2015). Overexpression of KLC2 due to a homozygous deletion in the non-coding region causes SPOAN syndrome. *Hum. Mol. Genet.* 24, 6877–6885. doi: 10.1093/hmg/ddv388
- Mendoza-Ferreira, N., Coutelier, M., Janzen, E., Hosseinibarkoobe, S., Löhr, H., Schneider, S., et al. (2018). Biallelic CHP1 mutation causes human autosomal recessive ataxia by impairing NHE1 function. *Neurol. Genet.* 4:e209. doi: 10.1212/NXG.0000000000000209
- Mitsuhashi, S., and Matsumoto, N. (2020). Long-read sequencing for rare human genetic diseases. *J. Hum. Genet.* 65, 11–19. doi: 10.1038/s10038-019-0671-8
- Mollet, J., Delahodde, A., Serre, V., Chretien, D., Schlemmer, D., Lombes, A., et al. (2008). CABP1 gene mutations cause ubiquinone deficiency with cerebellar ataxia and seizures. *Am. J. Hum. Genet.* 82, 623–630. doi: 10.1016/j.ajhg.2007.12.022
- Morais, S., Raymond, L., Mairey, M., Coutinho, P., Brandão, E., Ribeiro, P., et al. (2017). Massive sequencing of 70 genes reveals a myriad of missing genes or mechanisms to be uncovered in hereditary spastic paraplegias. *Eur. J. Hum. Genet.* 25, 1217–1228. doi: 10.1038/ejhg.2017.124
- Moreno-Cabrera, J. M., del Valle, J., Castellanos, E., Feliubadaló, L., Pineda, M., Brunet, J., et al. (2020). Evaluation of CNV detection tools for NGS panel data in genetic diagnostics. *Eur. J. Hum. Genet.* 28, 1645–1655. doi: 10.1038/s41431-020-0675-z
- Moreno-De-Luca, A., Helmers, S. L., Mao, H., Burns, T. G., Melton, A. M. A., Schmidt, K. R., et al. (2011). Adaptor protein complex-4 (AP-4) deficiency causes a novel autosomal recessive cerebral palsy syndrome with microcephaly and intellectual disability. *J. Med. Genet.* 48, 141–144. doi: 10.1136/jmg.2010.082263
- Moresco, E. M. Y., Li, X., and Beutler, B. (2013). Going forward with genetics: recent technological advances and forward genetics in mice. *Am. J. Pathol.* 182, 1462–1473. doi: 10.1016/j.ajpath.2013.02.002
- Ngo, K. J., Rexach, J. E., Lee, H., Petty, L. E., Perlman, S., Valera, J. M., et al. (2020). A diagnostic ceiling for exome sequencing in cerebellar ataxia and related neurological disorders. *Hum. Mutat.* 41, 487–501. doi: 10.1002/humu.23946
- Nguengang Wakap, S., Lambert, D. M., Olry, A., Rodwell, C., Gueydan, C., Lanneau, V., et al. (2020). Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *Eur. J. Hum. Genet.* 28, 165–173. doi: 10.1038/s41431-019-0508-0
- Novarino, G., Fenstermaker, A. G., Zaki, M. S., Hofree, M., Silhavy, J. L., Heiberg, A. D., et al. (2014). Exome sequencing links corticospinal motor neuron disease to common neurodegenerative disorders. *Science* 343, 506–511. doi: 10.1126/science.1247363
- Omenn, G. S., Lane, L., Lundberg, E. K., Overall, C. M., and Deutsch, E. W. (2017). Progress on the HUPO draft human proteome: 2017 metrics of the human proteome project. *J. Proteome Res.* 16, 4281–4287. doi: 10.1021/acs.jproteome.7b00375
- Orthmann-Murphy, J. L., Salsano, E., Abrams, C. K., Bizzi, A., Uziel, G., Freidin, M. M., et al. (2009). Hereditary spastic paraplegia is a novel phenotype for GJA12/GJC2 mutations. *Brain* 132, 426–438. doi: 10.1093/brain/awn328
- Ouahchi, K., Arita, M., Kayden, H., Hentati, F., Hamida, M. B., Sokol, R., et al. (1995). Ataxia with isolated vitamin E deficiency is caused by mutations in the  $\alpha$ -tocopherol transfer protein. *Nat. Genet.* 9, 141–145. doi: 10.1038/ng0295-141
- Palermo, A., Huan, T., Rinehart, D., Rinschen, M. M., Li, S., O'Donnell, V. B., et al. (2020). Cloud-based archived metabolomics data: a resource for in-source fragmentation/annotation, meta-analysis and systems biology. *Anal. Sci. Adv.* 1, 70–80. doi: 10.1002/ansa.202000042
- Parodi, L., Coarelli, G., Stevanin, G., Brice, A., and Durr, A. (2018). Hereditary ataxias and paraparesias: clinical and genetic update. *Curr. Opin. Neurol.* 31, 462–471. doi: 10.1097/WCO.0000000000000585
- Paulson, H. (2018). “Repeat expansion diseases” in *Handbook of clinical neurology*. eds. D. H. Geschwind, H. L. Paulson and C. Klein (Elsevier B.V.), 105–123.
- Pena, L. D. M., Jiang, Y. H., Schoch, K., Spillmann, R. C., Walley, N., Stong, N., et al. (2018). Looking beyond the exome: a phenotype-first approach to molecular diagnostic resolution in rare and undiagnosed diseases. *Genet. Med.* 20, 464–469. doi: 10.1038/gim.2017.128
- Pengelly, R. J., Ward, D., Hunt, D., Mattocks, C., and Ennis, S. (2020). Comparison of Mendeliome exome capture kits for use in clinical diagnostics. *Sci. Rep.* 10:3235. doi: 10.1038/s41598-020-60215-y

- Perez-Lloret, S., van de Warrenburg, B., Rossi, M., Rodríguez-Blázquez, C., Zesiewicz, T., Saute, J. A. M., et al. (2020). Assessment of ataxia rating scales and cerebellar functional tests: critique and recommendations. *Mov. Disord.* 36, 283–297. doi: 10.1002/mds.28313
- Platt, C. D., Zaman, F., Bainter, W., Stafstrom, K., Almutairi, A., Reigle, M., et al. (2020). Efficacy and economics of targeted panel versus whole-exome sequencing in 878 patients with suspected primary immunodeficiency. *J. Allergy Clin. Immunol.* 147, 723–726. doi: 10.1016/j.jaci.2020.08.022
- Pode-Shakked, B., Finezilber, Y., Levi, Y., Liber, S., Fleischer, N., Greenbaum, L., et al. (2020). Shared facial phenotype of patients with mucopolidosis type IV: a clinical observation reaffirmed by next generation phenotyping. *Eur. J. Med. Genet.* 63:103927. doi: 10.1016/j.ejmg.2020.103927
- Posey, J. E. (2019). Genome sequencing and implications for rare disorders. *Orphanet J. Rare Dis.* 14:153. doi: 10.1186/s13023-019-1127-0
- Rexach, J., Lee, H., Martinez-Agosto, J. A., Németh, A. H., and Fogel, B. L. (2019). Clinical application of next-generation sequencing to the practice of neurology. *Lancet Neurol.* 18, 492–503. doi: 10.1016/S1474-4422(19)30033-X
- Rodriguez, C. M., and Todd, P. K. (2019). New pathologic mechanisms in nucleotide repeat expansion disorders. *Neurobiol. Dis.* 130:104515. doi: 10.1016/j.nbd.2019.104515
- Roos, A., Thompson, R., Horvath, R., Lochmüller, H., and Sickmann, A. (2018). Intersection of proteomics and genomics to “solve the unsolved” in rare disorders such as neurodegenerative and neuromuscular diseases. *Proteomics Clin. Appl.* 12:1700073. doi: 10.1002/preca.201700073
- Rouleau, G. A., Ross, J. P., and Dion, P. A. (2020). Exome sequencing in genetic disease: recent advances and considerations. *F1000Research* 9:336. doi: 10.12688/f1000research.19444.1
- Ruano, L., Melo, C., Silva, M. C., and Coutinho, P. (2014). The global epidemiology of hereditary ataxia and spastic paraplegia: a systematic review of prevalence studies. *Neuroepidemiology* 42, 174–183. doi: 10.1159/000358801
- Sagnelli, A., Magri, S., Farina, L., Chiapparini, L., Marotta, G., Tonduti, D., et al. (2016). Early-onset progressive spastic paraplegia caused by a novel TUBB4A mutation: brain MRI and FDG-PET findings. *J. Neurol.* 263, 591–593. doi: 10.1007/s00415-016-8020-8
- Sánchez-Ferrero, E., Coto, E., Beetz, C., Gámez, J., Corao, A., Díaz, M., et al. (2013). 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. doi: 10.1111/j.1399-0004.2012.01896.x
- Sarzi, E., Goffart, S., Serre, V., Chrétien, D., Slama, A., Munnich, A., et al. (2007). Twinkle helicase (PEO1) gene mutation causes mitochondrial DNA depletion. *Ann. Neurol.* 62, 579–587. doi: 10.1002/ana.21207
- Segal, M. M., George, R., Waltman, P., El-Hattab, A. W., James, K. N., Stanley, V., et al. (2020). Clinician-centric diagnosis of rare genetic diseases: performance of a gene pertinence metric in decision support for clinicians. *Orphanet J. Rare Dis.* 15:191. doi: 10.1186/s13023-020-01461-1
- Seixas, A. I., Loureiro, J. R., Costa, C., Ordóñez-Ugalde, A., Marcelino, H., Oliveira, C. L., et al. (2017). A pentanucleotide ATTTC repeat insertion in the non-coding region of DAB1, mapping to SCA37, causes spinocerebellar ataxia. *Am. J. Hum. Genet.* 101, 87–103. doi: 10.1016/j.ajhg.2017.06.007
- Sharon, R., Martins, S., Costa, I. P. D., Zaltzman, R., Amorim, A., Sequeiros, J., et al. (2019). Yemenite-Jewish families with Machado-Joseph disease (MJD/SCA3) share a recent common ancestor. *Eur. J. Hum. Genet.* 27, 1731–1737. doi: 10.1038/s41431-019-0449-7
- Shribman, S., Reid, E., Crosby, A. H., Houlden, H., and Warner, T. T. (2019). Hereditary spastic paraplegia: from diagnosis to emerging therapeutic approaches. *Lancet Neurol.* 18, 1136–1146. doi: 10.1016/S1474-4422(19)30235-2
- Simons, C., Wolf, N. I., McNeil, N., Caldovic, L., Devaney, J. M., Takanohashi, A., et al. (2013). A de novo mutation in the  $\beta$ -tubulin gene TUBB4A results in the leukoencephalopathy hypomyelination with atrophy of the basal ganglia and cerebellum. *Am. J. Hum. Genet.* 92, 767–773. doi: 10.1016/j.ajhg.2013.03.018
- Singh, J. (2020). Harnessing the power of genetics: fast forward genetics in *Caenorhabditis elegans*. *Mol. Gen. Genomics.* 296, 1–20. doi: 10.1007/s00438-020-01721-6
- Ślabicki, M., Theis, M., Krastev, D. B., Samsonov, S., Mundwiler, E., Junqueira, M., et al. (2010). A genome-scale DNA repair RNAi screen identifies SPG48 as a novel gene associated with hereditary spastic paraplegia. *PLoS Biol.* 8:e1000408. doi: 10.1371/journal.pbio.1000408
- Sobreira, N., Schiettecatte, F., Valle, D., and Hamosh, A. (2015). GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum. Mutat.* 36, 928–930. doi: 10.1002/humu.22844
- Srivastava, S., Love-Nichols, J. A., Dies, K. A., Ledbetter, D. H., Martin, C. L., Chung, W. K., et al. (2019). Meta-analysis and multidisciplinary consensus statement: exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders. *Genet. Med.* 21, 2413–2421. doi: 10.1038/s41436-019-0554-6
- Subramanian, I., Verma, S., Kumar, S., Jere, A., and Anamika, K. (2020). Multi-omics data integration, interpretation, and its application. *Bioinform. Biol. Insights* 14:1177932219899051. doi: 10.1177/1177932219899051
- Sud, M., Fahy, E., Cotter, D., Azam, K., Vadivelu, I., Burant, C., et al. (2016). Metabolomics workbench: an international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Res.* 44, D463–D470. doi: 10.1093/nar/gkv1042
- Suwinski, P., Ong, C. K., Ling, M. H. T., Poh, Y. M., Khan, A. M., and Ong, H. S. (2019). Advancing personalized medicine through the application of whole exome sequencing and big data analytics. *Front. Genet.* 10:49. doi: 10.3389/fgene.2019.00049
- Szynofzik, M., and Schüle, R. (2017). Overcoming the divide between ataxias and spastic paraplegias: shared phenotypes, genes, and pathways. *Mov. Disord.* 32, 332–345. doi: 10.1002/mds.26944
- Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., and Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* 20, 467–484. doi: 10.1038/s41576-019-0127-1
- Thuriot, F., Buote, C., Gravel, E., Chénier, S., Désilets, V., Maranda, B., et al. (2018). Clinical validity of phenotype-driven analysis software PhenoVar as a diagnostic aid for clinical geneticists in the interpretation of whole-exome sequencing data. *Genet. Med.* 20, 942–949. doi: 10.1038/gim.2017.239
- Trottier, Y., Lutz, Y., Stevanin, G., Imbert, G., Devys, D., Cancel, G., et al. (1995). Polyglutamine expansion as a pathological epitope in huntington's disease and four dominant cerebellar ataxias. *Nature* 378, 403–406. doi: 10.1038/378403a0
- Trouillas, P., Takayanagi, T., Hallett, M., Currier, R. D., Subramony, S. H., Wessel, K., et al. (1997). International cooperative ataxia rating scale for pharmacological assessment of the cerebellar syndrome. *J. Neurol. Sci.* 145, 205–211. doi: 10.1016/S0022-510X(96)00231-6
- Utine, G. E., Haliloğlu, G., Salanci, B., Çetinkaya, A., Kiper, P. Ö., Alanay, Y., et al. (2013). A homozygous deletion in GRID2 causes a human phenotype with cerebellar ataxia and atrophy. *J. Child Neurol.* 28, 926–932. doi: 10.1177/0883073813484967
- van der Knaap, M. S., Leegwater, P. A. J., Könst, A. A. M., Visser, A., Naidu, S., Oudejans, C. B. M., et al. (2002). Mutations in each of the five subunits of translation initiation factor eIF2B can cause leukoencephalopathy with vanishing white matter. *Ann. Neurol.* 51, 264–270. doi: 10.1002/ana.10112
- Wagner, M., Osborn, D. P. S., Gehweiler, I., Nagel, M., Ulmer, U., Bakhtiar, S., et al. (2019). Bi-allelic variants in RNF170 are associated with hereditary spastic paraplegia. *Nat. Commun.* 10:4790. doi: 10.1038/s41467-019-12620-9
- Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., et al. (2016). Sharing and community curation of mass spectrometry data with global natural products social molecular networking. *Nat. Biotechnol.* 34, 828–837. doi: 10.1038/nbt.3597
- Wang, D., Eraslan, B., Wieland, T., Hallström, B., Hopf, T., Zolg, D. P., et al. (2019). A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol. Syst. Biol.* 15:e8503. doi: 10.15252/msb.20188503
- Wenger, A. M., Peluso, P., Rowell, W. J., Chang, P. C., Hall, R. J., Concepcion, G. T., et al. (2019). Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* 37, 1155–1162. doi: 10.1038/s41587-019-0217-9
- Winterthun, S., Ferrari, G., He, L., Taylor, R. W., Zeviani, M., Turnbull, D. M., et al. (2005). Autosomal recessive mitochondrial ataxic syndrome due to mitochondrial polymerase  $\gamma$  mutations. *Neurology* 64, 1204–1208. doi: 10.1212/01.WNL.0000156516.77696.5A
- Wishart, D. S. (2016). Emerging applications of metabolomics in drug discovery and precision medicine. *Nat. Rev. Drug Discov.* 15, 473–484. doi: 10.1038/nrd.2016.32
- Wright, C. F., FitzPatrick, D. R., and Firth, H. V. (2018). Paediatric genomics: diagnosing rare disease in children. *Nat. Rev. Genet.* 19, 253–268. doi: 10.1038/nrg.2017.116
- Yang, Z., Shi, C., Zhou, L., Li, Y., Yang, J., Liu, Y., et al. (2019). Metabolic profiling reveals biochemical pathways and potential biomarkers of spinocerebellar ataxia 3. *Front. Mol. Neurosci.* 12:159. doi: 10.3389/fnmol.2019.00159

- Yépez, V. A., Mertes, C., Müller, M. F., Klaproth-Andrade, D., Wachutka, L., Frésard, L., et al. (2021). Detection of aberrant gene expression events in RNA sequencing data. *Nat. Protoc.* 16, 1276–1296. doi: 10.1038/s41596-020-00462-5
- Ylikallio, E., Kim, D., Isohanni, P., Auranen, M., Kim, E., Lönnqvist, T., et al. (2015). Dominant transmission of de novo KIF1A motor domain variant underlying pure spastic paraplegia. *Eur. J. Hum. Genet.* 23, 1427–1430. doi: 10.1038/ejhg.2014.297
- Zanni, G., Cali, T., Kalscheuer, V. M., Ottolini, D., Barresi, S., Lebrun, N., et al. (2012). Mutation of plasma membrane Ca<sup>2+</sup> ATPase isoform 3 in a family with X-linked congenital cerebellar ataxia impairs Ca<sup>2+</sup> homeostasis. *Proc. Natl. Acad. Sci. U. S. A.* 109, 14514–14519. doi: 10.1073/pnas.1207488109
- Zhao, M., Havrilla, J. M., Fang, L., Chen, Y., Peng, J., Liu, C., et al. (2020). Phen2Gene: rapid phenotype-driven gene prioritization for rare diseases. *NAR Genomics Bioinform.* 2:lqaa032. doi: 10.1093/nargab/lqaa032
- Zivony-Elboum, Y., Westbroek, W., Kfir, N., Savitzki, D., Shoval, Y., Bloom, A., et al. (2012). A founder mutation in Vps37A causes autosomal recessive complex hereditary spastic paraparesis. *J. Med. Genet.* 49, 462–472. doi: 10.1136/jmedgenet-2012-100742

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Yahia and Stevanin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.