KU Leuven
Biomedical Sciences Group
Faculty of Medicine
Vlaams Instituut voor Biotechnologie (VIB)
Center for Brain & Disease Research (CBD)
Laboratory of Neurobiology



GENETIC FACTORS DETERMINING VULNERABILITY OF NEURONS INVOLVED IN MOTOR CONTROL

Sien Van Daele

Jury:

Supervisor:	Prof. Dr. Philip Van Damme
Co-supervisor:	Prof. Dr. Kristl Claeys
	Dr. Valérie Race
Chair examining committee:	Prof. Dr. Marc Fransen
Chair public defense:	Prof. Dr. Patrick Dupont
Jury members	Prof. Dr. An Goris
	Prof. Dr. Joris Vermeesch
Jury members:	Prof. Dr. Alexandra Durr
	Prof. Dr. Matthis Synofzik

Dissertation presented in partial fulfilment of the requirements for the degree of Doctor in Biomedical Sciences

02/05/2022

Table of Contents

Table of Contents 3
Acknowledgements 5
Abbreviations 11
Chapter 1: Introduction 15
Disorders affecting the motor system 15
The motor system 15
Amyotrophic lateral sclerosis 16
Hereditary ataxia 20
Hereditary spastic paraplegia 25
Heterogeneity and overlap in disorders affecting the motor system 26
Phenotypic heterogeneity and overlap 26
Genetic heterogeneity and overlap 27
Vulnerability of the motor system 28
Next generation sequencing 29
A revolutionary approach 29
Advantages and limitations 31
MPS in disorders affecting the motor system 33
Chapter 2: Objectives 35
Chapter 3: Genetic variability in known ALS genes in sporadic ALS patients 37
Abstract 37
Introduction 38
Methods 39
Results 42
Discussion 50
References 53
Supplementary data 56
Chapter 4: Genetic yield and phenotype-genotype correlations in a large ataxia-HSP cohort in Belgium 83
Abstract 83
Introduction 84
Methods 85
Results 86
Discussion 103
Conclusion 106
References 106
Supplementary data 110
Chapter 5: Diagnostic yield of testing for *RFC1* repeat expansions in patients with unexplained adult-onset cerebellar ataxia 119
Introduction 120
Patients and methods 120
Results 121
Discussion 123
References 123

Chapter 6: *RNF170* mutation causes autosomal dominant sensory ataxia with variable pyramidal involvement 125
Abstract 126
Introduction 127
Methods 127
Results 128
Discussion 129
References 132
Supplementary data 133
Chapter 7: General discussion and perspectives 137
MPS and gene panel analysis in disorders affecting the motor system 137
Yield and genotype-phenotype correlations 137
What with unexplained patients? 141
Complex genetic diseases 143
Overlap and selective vulnerability of the motor system 145
Future perspectives 146
Thesis abstract 147
References related to introduction and discussion 149
Scientific acknowledgements, personal contributions and conflicts of interest 155
Curriculum vitae 157

Acknowledgements

A thesis is never written alone. During the 4,5 years working on this PhD project, many colleagues, friends and family members stood by my side, before as well as behind the scenes. Writing this acknowledgements section may therefore be one of the hardest parts of all, since it is impossible to do right by all the persons that have guided me along the way. The list that follows is thus by definition non-exhaustive, but however, I will give it a try.

First of all, I want to thank my promotor, **Prof. Dr. Philip Van Damme**, for his expertise, encouragement and patience. Thank you for always being there and making the time when I needed advice. I am truly honored that you wanted to guide me during this journey.

I also want to thank **Prof. Dr. Kristl Claeys** and **Dr. Valérie Race**, my co-promotors, for their feedback during these 4 years. Your expertise during the neurogenetics meetings have learned me a lot.

Next I want to thank **Prof. Dr. Ludo Van Den Bosch** and **Prof. Dr. Robin Lemmens** to welcome me in the lab of Neurobiology. Ludo, going somewhere with you is never dull. Thank you for the nice chats now and then, at least before corona threw a spanner in the works…

My gratitude also goes to the internal and external jury members, **Prof. Dr. Joris Vermeesch, Prof. Dr. An Goris, Prof. Dr. Matthis Synofzik and Prof. Dr. Alexandra Durr**, who took the time and effort to read through my thesis and formulate their valuable comments and feedback.

Four and a half years ago, I arrived in the lab of Neurobiology. Although I have always been a bit 'the odd man out' because I did not work with cells, flies or zebrafish, I have always felt welcome. First of all, I am very grateful to **Matthieu**, the only other person in the lab staying away from the more typical lab equipment, for introducing me in the complex world of data analysis, genetics and bio-informatics. I suppose it has taken a lot of patience from you to train an IT-dummy like me. I think it's safe to say that you were the operating system of my thesis; without you I would never have gotten where I am now. I wish you all the luck with GenAdvice, I'm sure you will do great !

When I arrived in the lab, I was very lucky with the desk that was allocated to me. I was seated in the comfortable environment that was later renamed into 'Het Hoekske' and where, besides hard work, also fun and pleasure were number one priorities. No topic was taboo, from photography advice to wedding recommendations. And I'm absolutely positive that we had the nicest Christmas

decorations of the lab! Therefore, a very big thank you to the original line-up of ‘Het Hoekske’; **Mathias**, **Jolien**, and **Wenting**, and the later assets **Elisabeth**, **Joni**, **Antina** and **Raheem** ! You made sure that being at work was never dull !

Stijn, I think you were the most common visitor of ‘het Hoekske’ and it was always a pleasure when you came by. Thanks for the many talks, I love how you always see the positive side of everything. Together with Mathias and **Bart**, who was always in for a joke as well (but not for sweatpants apparently, so we learned), you also hosted the ‘Mr Friday’ eating events, something I could look forward to for the entire week !

Donya, I got to know you on the PhD class when we both started. I love your direct but very sweet and considerate attitude and I have always loved chatting with you! I wish you all the best in the UK and hope we can meet again soon.

Tom, the other UK (but also partly Belgian) guy, I enjoyed your dry English humor ! I learned a lot of English expressions from you (mum’s the word!) and hope I taught you some Belgian ones as well.

Nicole, you were always there (and for the most part literally, since you had the desk across from me). You were always ready to help and we had nice talks as well. Thanks for lending us the flowerpot for our little Christmas tree! Since we live very close now, I hope to see you regularly in the future.

Rob, you were also a constant factor during my PhD years. I admire your passion for both work and pleasure, I think ‘work hard play hard’ fits perfect for you. Stay as you are!

Laura F., I love your Italian vibes and it was always fun being with you. Your photography skills were amazing, I hope you keep on exploring them further. I wish you all the best in your next challenge !

Also **Wendy** and **Annet** are for me ‘one with the lab of Neurobiology’. It was always nice walking in to you and having a little chat !

Pegah, as a fellow neurologist in training, we had something else in common besides our passion for research. It was always a pleasure talking with you, and sharing our experiences was every time enriching. I hope we will continue to see each other. I wish you all the best with your project, I admire your persistence and am sure it will turn out great !

Joke T., Joke DV, Nikita and **Fouke**, we had the same roots in the Neurology department of the hospital. I admire your work attitudes and enjoyed all the chats we had along the way!

Arun, you work hard but always take the time for a kind word. I'm very grateful for your help and patience with the 3D plot!

Valérie, **Adrià**, **Laura R.**, **Elke, Ann, Begga, Caroline, Lindsay, Katarina, Jimmy**, **Paraskevi**, **Séraphina, Silke, Christine, Kristel, Alessio** and **Yvonne**, although we didn't really work together, it was always a pleasure to see you in the corridors or on lab-related activities. I enjoyed the chats that we had !

Due to the well known virus and the final stretch of my PhD, I haven't been a lot in the lab lately… Since the lab is a dynamic environment, many new colleagues have joined since then. **Thibaut**, **Marjam**, **Marie-Lou**, **Dufie**, **Melissa**, **Tessa**, **Kara**, **Silke**, **Lisha**, **Evelien**, **Jordi**, **Ilse**, **Clara**, **Adriana**, **Matilde**, **Bennie** and **Daina**, I hope you all had a good start and I wish you all the best during your time here!

Us PhD students would never survive without someone to look after us and to answer our many logistic questions. **Lien**, **Elisa**, **Annelies** and **Nuray**, thank you for always being there and doing such an amazing job !

My time in the lab is intrinsically linked to my residency in the Neurology department of UZ Leuven. I want to thank everyone there for the support, and especially **Ann D'Hondt** and the neuromuscular center for the (study) coordination, and **Stefanie** and **Cindy** for answering all my questions with much patience.

Some colleagues have become real friends, and their support and listening ear have been invaluable through these years. **Aline**, **Anke**, **Annelies**, **Caroline**, **Ide**, **Katrien**, **Lynn**, **Marijke** and **Sarah**: thanks for all the fantastic moments during and besides work ! A special shout-out goes to **Ide**, who I consider 'the godmother' of my PhD, giving me lots of advice and being my soundboard. After she left, this role was gloriously fulfilled by **Aline**, a true friend who is always there when you need her.

Since my PhD is not only rooted in neurology, but also in genetics, I worked together with the center of human genetics (CME). I want to thank everyone from the neurogenetics team there; **Valérie**, **Lies**, **Amélie**, **Prof. Dr. Sascha Vermeer** and **Prof. Dr. Hilde Van Esch** for the nice collaboration. You strengthened my fascination for genetics and I have learned a lot from our monthly meetings.

As my fascination for genetics grew, I sought for a way to further explore this interest and I was lucky enough to be welcomed by the CME for a further education, an opportunity for which I am most grateful. I want to thank everyone there for their support and willingness to 'take me in'. Especially my newly found colleagues **Laurens**, **Marie**, **Mathilde**, **Michiel**, **Miel**, **Mio**, **Nathalie** have been most welcoming and supporting, making me feel confident that the next years will be at least as delightful as the previous ones.

As I wrote in the beginning of this section, honoring anyone who was valuable for me these last 4,5 years is an impossible assignment, especially the ones behind the scenes. Therefore, I won't take on that challenge, but there are some persons I specifically want to thank anyway.

First of all, **Margot**, as a fellow PhD student (and post-doc) you knew exactly what the highs and lows of the journey were. Thank you for being there when I needed it, listening to me and supporting me. I wish you all the best in the US (but secretly I also hope you come back soon).

And then, the true pillars that I lean on, my family that I can always count on. **Mama**, whatever the question or the problem, you are always there to listen and to help. Thank you for all your support, both 'mental' and practical, I don't know how I would have done it without you. **Papa**, thank you for always being there when I need advice and making the time when I need your help. Your commitment and persistence have always been an example for me. **Ruth**, our PhD trajectories coincided (you were a bit faster than me) and you understood better than anyone what the issues could be. Thanks for your help and proofreading in the end-stage, with your critical, well-trained eye as always. **Nel**, I have deep respect for your hard work ethos, which you in some way manage to combine with a rich social life. Although you live somewhat further from the rest of us, you still make the time to help when needed and I'm truly grateful for that. **Tuur,** my favorite brother, you also know the peaks and downs from PhD life. You work hard and you have a heart of gold, treasure it ! You are there when I need you, thank you for that! Also a big thanks to **my grandparents, Iris, Dimitri**, **Els, Lieve, Guido** and **Karel** for their love and support!

And finally, last but absolutely not least, I want to thank **Kris**, my one and only, my soulmate. Thank you for ALWAYS being there for me, for listening to my small and big worries and for always having my back. You supported me in every decision that I made, even when it was not always the easiest one. Without you, none of this would have been possible. I enjoyed that I could catch a glimpse of your area of interest by working with databases and I really am grateful for your help and teaching sessions with that.

And to end, my sweet **Lexi**, you are the most beautiful real-life application of genetics that I could ever have imagined. Your sparkling presence energizes me every day (despite the sometimes short nights). You give me strength and courage and help me to put things in perspective.

A big thank you to you all!

Sien

Abbreviations

ACMG	American College of Medical Genetics and Genomics
AD	autosomal dominant
ADCA	autosomal dominant cerebellar ataxia
ADD	attention deficit disorder
Adj	adjusted
ADSA	autosomal dominant sensory ataxia
ALS	amyotrophic lateral sclerosis
amb	ambiguous
AO	age at onset
AOA1	ataxia with oculomotor apraxia type I
AOA2	ataxia with oculomotor apraxia type 2
AR	autosomal recessive
ARCA	autosomal recessive cerebellar ataxia
ARSACS	autosomal recesive spastic ataxia of charlevoix-saguenay
AT	ataxia
AVS	aortic valve stenosis
B	benign
bp	base pairs
C	constipation
CANVAS	cerebellar ataxia, neuropathy and vestibular areflexia syndrome
CB	cerebellar
CCD1	coiled-coil domain 1
cDNA	complementary DNA
CI	confidence interval
CMAP	compound muscle action potential
CMT	Charcot-Marie-Tooth disease
CNV	copy number variation
CT	cerebrotendinous xanthomatosis
D	dysarthria
dHMN	distal hereditary motor neuropathy
DM	demyelinating
DRPLA	dentatorubral pallidoluysian atrophy
dup	Duplication
EA2	episodic ataxia type 2
EMG	electromyography
Est	estimated
F	female
fALS	familial amyotrophic lateral sclerosis
FDA	food and drug administration
FDI	first dorsal interosseous
FEWDON	finger extension weakness and downbeat nystagmus motor neuron disease syndrome
FTD	frontotemporal dementia
FXTAS	fragile X-associated tremor/ataxia syndrome
GA	gait ataxia
GP	globus pallidus
GQ	genotype quality
GRCh37	genome reference consortium human build 37
GWAS	genome-wide association studies
HA	hereditary ataxia
Hom	homozygous

HoR	hyporeflexia
HR	hyperreflexia
HSP	hereditary spastic paraplegia
HT	hypertonia
HTZ	heterozygous
I	impotence
Inf	infinity
inv	inversion
IP	inheritance pattern
LA	limb ataxia
LB	likely benign
LH	linear hypointensities
LL	lower limbs
LOF	loss-of-function
LP	likely pathogenic
LV	list of variants
M	male
MAF	minor allele frequency
MASA	mental retardation, aphasia, shuffling gait and adducted thumbs
max	maximum
MI	mitral insufficiency
min	minimum
MIS	missense
MPS	massive parallel sequencing
MRI	magnetic resonance imaging
MSA	multiple system atrophy
N	no
N	number
NA	not available
NCS	nerve conduction study
NGS	next generation sequencing
NMHP	neuromuscular homepage
NP	not performed
Ny	nystagmus
OA	oculomotor apraxia
OD	ocular dysmetria
OH	orthostatic hypotension
OR	odds ratio
P	pathogenic
Pe	peripheral
PM	point mutation
PT	patient
Py	pyramidal
RE	repeat expansion
RL	repeat length
RRM	RNA recognition motif domain
RT	retinal telangiectasia
S	sporadic
sALS	sporadic amyotrophic lateral sclerosis
SBMA	spinobulbar muscular atrophy
SCA	spinocerebellar ataxia
SCAR	autosomal recessive spinocerebellar ataxia
SE	saccadic eye movements
SM	sensorimotor

SNAP	sensory nerve action potential
SPAX	spastic ataxia
SPG	spastic paraplegia
SS	slow saccades
SSEP	somatosensory evoked potentials
SV	sequencing variant
TRUNC	truncating
TS	targeted sequencing
U	urinary dysfunction
UL	upper limbs
ULD	ubiquitin-like domain
UTR	untranslated region
VI	variant of interest
VUS	variant of unknown significance
VUS+	variant of unknown significance with pathogenic evidence
WES	whole exome sequencing
WGS	whole genome sequencing
WML	white matter lesions
X	X-linked
XL	X-linked
Y	yes
Zyg	zygosity

Chapter 1: Introduction

Disorders affecting the motor system

The motor system

The first scientific reports about the human motor system date back to the 2nd century.[1] Given the extraordinarily refined movements that can be carried out by dancers, musicians and even by all humans in daily life, it should come as no surprise that the unravelment of this complex organized network has not yet been accomplished. Research is still ongoing by neuroscientists and other researchers every day. While it is obvious that the motor system controls voluntary movements, it is less clear what exactly constitutes this motor system. Evidently, the primary motor cortex, upper and lower motor neurons form the very core, generating and carrying motor signals through the corticospinal tract and the peripheral neurons to the muscles. However, many other structures are involved. Different circuits and feedback loops are involved and refine and control every motion to obtain the smooth and precise action that we often deem so naturally. Examples are other cortical areas such as the premotor cortex and supplementary motor cortex, the basal ganglia and even the sensory system. Also the cerebellum has an important role to play, ensuring a smooth coordination of ongoing movements and participating in motor planning.[2]

Sadly enough, this sophisticated machinery is not flawless and damage to one or more of its components can have major consequences. Although many mechanisms can possibly damage the motor system, e.g. ischemia (stroke), infection, inflammation and tumors; in this thesis we focus on three neurodegenerative disorders primarily affecting the motor system (Fig. 1). Amyotrophic lateral sclerosis (ALS) is probably best known, and is characterized by combined loss of upper and lower motor neurons. In hereditary ataxia (HA), the typical sites of degeneration are the cerebellum and its connections resulting in progressive prominent ataxia. Finally, hereditary spastic paraplegia (HSP) primarily affects the upper motor neurons causing progressive spasticity in mainly the lower limbs. All these disorders share three important features: there often is a genetic origin, the neurons/axons that make up the motor system are degenerating as part of the clinical spectrum and no cure exists.[3-5] Although they present differently at the clinical level in typical cases, evidence is emerging that overlap at the clinical and genetic level exists between the different disorders, as is further discussed below.

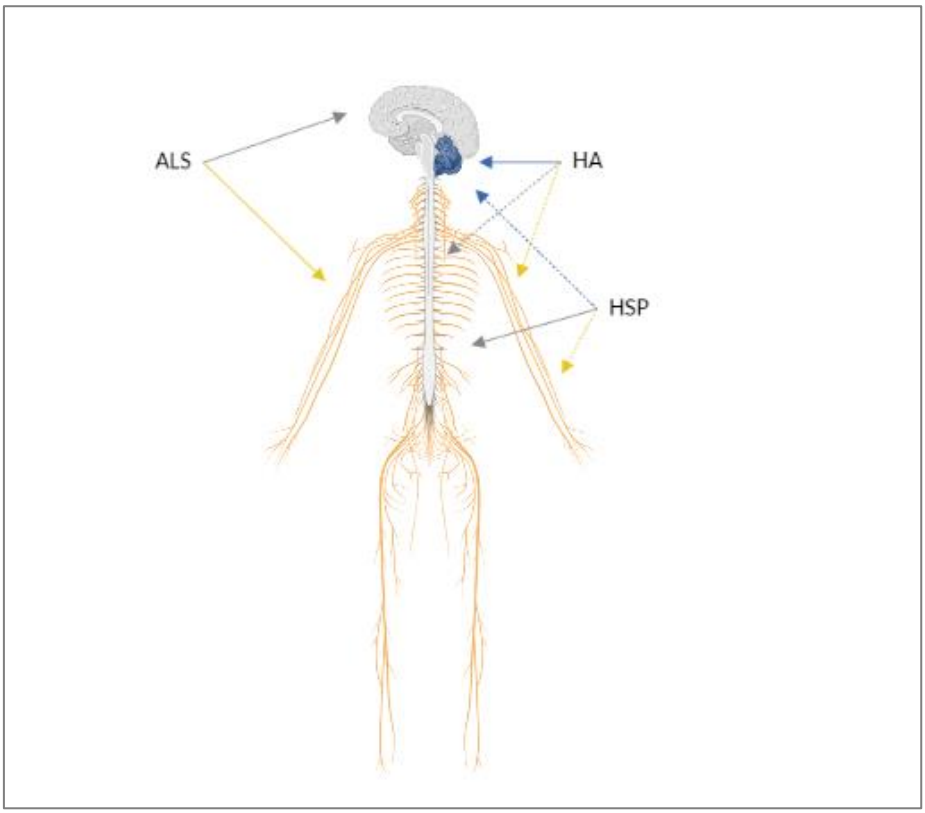


Figure 1: Disorders affecting the motor system.
ALS, HA and HSP all affect parts of the motor system, namely the upper motor neuron (UMN) system (depicted in grey), the lower motor neurons (LMN) (in yellow), and the cerebellum (in blue). ALS typically affects the UMNs and LMNs (full arrows). Rare subtypes of motor neuron disorders can also affect the cerebellar pathways (not shown). Hereditary ataxia is characterized by cerebellar dysfunction (full arrow), although the UMN and LMN can also be involved (dashed arrows). HSP affects the spinal UMNs (full arrow), but also the cerebellum and LMNs can be implicated (dashed arrows). *Created with BioRender.com.*

Abbreviations: ALS, amyotrophic lateral sclerosis; HA, hereditary ataxia; HSP, hereditary spastic paraplegia; LMN, lower motor neuron; UMN, upper motor neuron.

Amyotrophic lateral sclerosis

ALS is a degenerative motor neuron disorder characterized by upper and lower motor neuron degeneration, which leads to muscle weakness and other signs of motor dysfunction.[3] Next to the motor signs, up to 50% of the patients develop cognitive and/or behavioral dysfunction as part of a spectrum with frontotemporal dementia (FTD).[3] With an incidence of 2-3 cases per 100 000 individuals per year, ALS is considered a rare disease.[3] However, lifetime risk is as high as 1 in 400 and with the ageing population, the burden of motor neuron diseases is very likely to increase substantially in the near future.[6, 7] The disease is relentlessly progressive and most patients die between 2-5 years after disease onset.[8] Only two drugs, Riluzole and Edaravone, have been approved by the Food and Drug Administration (FDA) but their beneficial effect is only minimal.[9, 10] The management of ALS is therefore mainly limited to supportive care, preferentially by a multidisciplinary approach.[3] Novel drugs are in development, with hopeful expectations for antisense nucleotides (ASO's) among others.[11, 12]

Despite extensive research, the etiology of ALS has not been fully unraveled to date. Traditionally a distinction is made between cases with a familial history of ALS/FTD (fALS) representing 5-10% of the cases, and the remaining apparently sporadic patients (sALS).[13] Although there is no strict definition to date, it is commonly accepted that an ALS patient with one or more first- or second degree relatives with ALS can be classified as having fALS.[14] In these familial cases, pedigrees often show a (mainly autosomal dominant) Mendelian inheritance pattern which is suggestive for highly penetrant mutations.[15] Pathogenic monogenetic variants are found in ~70% of familial patients, most commonly in *C9orf72*, *SOD1*, *TARDPB* and *FUS* which are responsible for about 40%, 20%, 4% and 3% of familial cases, respectively, at least in European populations (Fig. 2).[16] In *C9orf72*, hexanucleotide intronic repeat expansions have been identified, while in the other genes, pathogenic sequencing variants are found.[16, 17] By contrast, sporadic ALS was -and sometimes still is- often regarded as non-genetic and seen as an 'environmental disease'. However, in ~10% of sporadic cases, mutations in the same genes as in fALS are found (Fig. 2).[18] Furthermore, based on twin-studies, heritability of sALS is estimated to be around 60%.[19] Family members of patients with sporadic ALS are shown to have an approximately eight times higher risk of developing ALS compared to the background population, although this only results in slightly increased absolute risks.[20] In fALS, reduced penetrance and genetic pleiotropy often characterize families, suggesting that other (genetic) factors might play a role.[3] As the lifetime risk of ALS is as high as 1/400, the chance of having a second family member with ALS just by chance is relatively high, especially in large families, meaning that these probands would wrongly be classified as having fALS.[14] Given all this and since patients with familial and sporadic ALS cannot be discriminated on a phenotypic basis, it seems that the distinction between 'familial' and 'sporadic' cases is an oversimplification.[21]

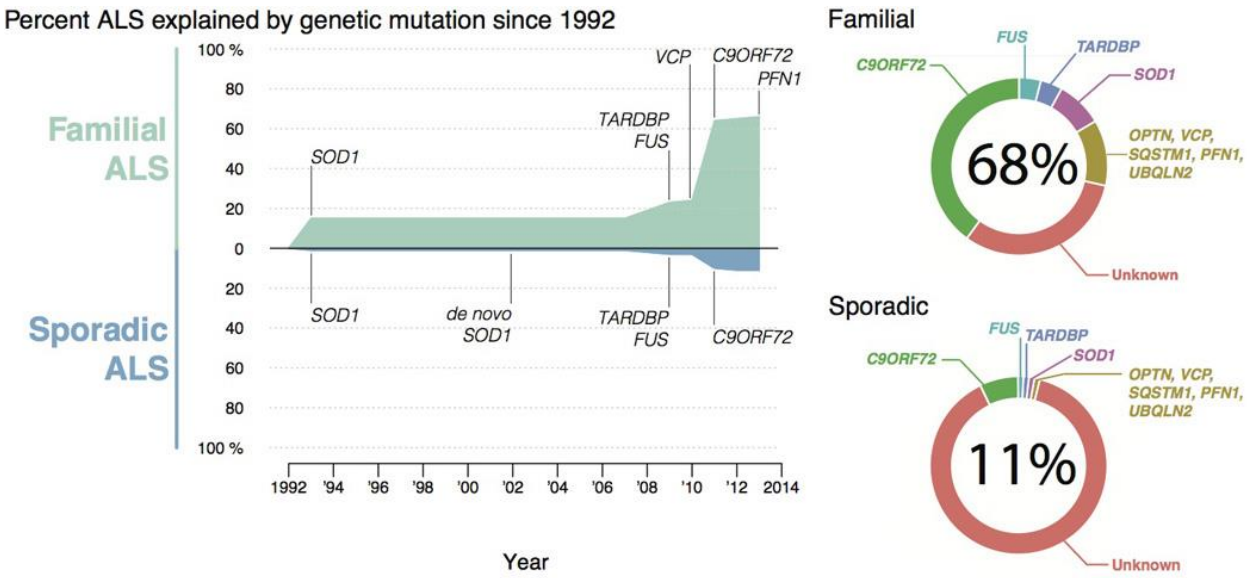


Figure 2: Timeline of gene discoveries in familial and sporadic ALS (left) and overview of genes harboring pathogenic variants in patients with familial and sporadic ALS (right).
LEFT: timeline: values on the y-axis represent the total proportion of ALS patients of European ancestry that are genetically explained at the corresponding year on the x-axis (top: familial ALS, bottom: sporadic ALS)
RIGHT: the number in the circle indicates how many patients (top: familial, bottom: sporadic) are genetically explained by single gene variants. The distribution in the circle plot indicates which proportion of explained patients has a variant in the corresponding gene.
Remark: The timeline in this figure runs until 2014. Since then, additional genes have been linked with ALS but overall values represented in the pie plot remain currently valid.
Reproduced from Renton et al. (2015); with permission.[16]

In the last years, more and more evidence has arisen for an oligogenic inheritance model for ALS. For example, it has been shown that the presence of multiple mutations in fALS was in excess of what was to be expected by chance.[22] Furthermore, the low parent-offspring and twin concordance rates are most consistent with models involving multiple variants rather than single highly penetrant variants.[23] With this oligogenic hypothesis in mind, we can look at familial and sporadic ALS as being part of a spectrum (Figure 2). On the one end, we find patients with single, highly penetrant mutations that lead to Mendelian inheritance and clear familial ALS. On the other end, there are patients with multiple less penetrant mutations together that constitute their risk on ALS. These patients are typically not part of families with Mendelian inheritance of ALS, although family members might be at slightly higher risk because of their genetic constitution.

All this is compatible with the multistep model that has been proposed for ALS, which postulates that overall six steps are needed to develop ALS.[24] For patients with large-effect Mendelian gene variants, these variants might already count for multiple steps together, therefore only needing few other genetic and/or environmental triggers to develop the disease.

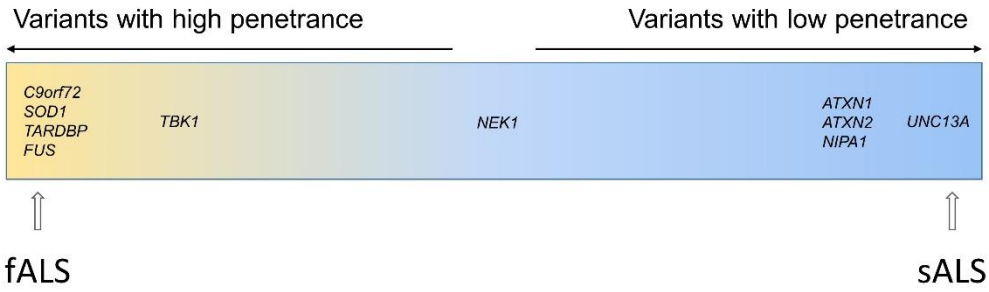


Figure 3: Spectrum of penetrance in ALS.
On the one end of the spectrum (left), patients carry one or few highly penetrant variants. These probands are typically part of families with clearly Mendelian inheritance patterns. On the other end of the spectrum, patients have variants with very low penetrance each, rather to be seen as risk factors, and multiple variants together are needed before these patients develop ALS. Some known ALS genes are plotted as example in the spectrum (positions not in concordance with their true penetrance ratios). In *C9orf72*, *ATXN1*, *ATXN2* and *NIPA1* (intermediate) repeat expansions are linked with ALS.[17, 25-27] In *SOD1*, *TARDBP*, *FUS*, *TBK1* and *NEK1* sequencing variants are found.[16, 28, 29] In *UNC13A* the CC-genotype of the polymorphism rs12608932 has been associated with ALS.[30]

Abbreviations: fALS, familial ALS; sALS, sporadic ALS.

Despite these insights, the exact missing heritability in (sporadic) ALS remains largely unexplained. During the last years, a lot of research has been done trying to find the missing genetic factors.[15, 31, 32] Originally, some genome-wide association studies (GWAS) identified several robust ALS risk loci and common variants with small odds ratios, suggesting that common variants played an important role in the development of ALS.[30] Risk factors most commonly acknowledged are *UNC13A* (homozygosity for the C-allele of SNP rs12608932; OR ~1.29), *NIPA1* repeat expansions (identified by means of a CNV association study; OR ~1.71), *ATXN2* repeat expansions (identified through a modifier screen in yeast for genes that affect TDP-43 toxicity; OR ~3.06-4.44) and *ATXN1* repeat expansions (identified in analogy with *ATNX2* repeat expansions; OR ~1.38).[25, 27, 31, 33-35] However, the top loci from a recent imputation and mixed-model association analysis could only explain 0.2% from the estimated 8.5% SNP-based heritability.[31] Therefore it was suggested that a distinct role is reserved for rare variants, which are not easily captured by GWAS studies. Other techniques such as massive parallel sequencing (MPS) are therefore increasingly implemented, as is described below. As a result from these and other various studies, a considerable number of variants and/or genes have been associated with ALS through a number of different genetic approaches. For example, recently pathogenic repeat expansions in *HTT*, usually associated with Huntington's disease, were discovered as a rare cause of FTD/ALS spectrum disease.[36] However, not for all of the identified genes the evidence is equally convincing and quite some findings failed to be replicated in subsequent studies.[23] Also on variant level, uncertainty often exists about the

strength and nature of association of a specific variant with ALS, even in genes generally accepted to be linked to ALS.

Finally, another problem in ALS research is that, until now, genetic research has mainly focused on familial ALS since the genetic component of the disease is most evident in these patients. Large studies describing genetic variation on sALS patients are currently lacking. However, a genetic characterization of all ALS patients becomes increasingly important now gene therapies for ALS are on the horizon.

Hereditary ataxia

Ataxia (Greek ‘a’ without, ‘taxis’ order) describes a lack of coordination of voluntary movements.[37] While the muscle strength can be completely normal per se, patients have trouble coordinating their muscle activity resulting in irregular and clumsy movements, balance problems, speech abnormalities, swallowing problems and eye movement abnormalities. The lesion is most often localized at the cerebellum or its connections such as the spinocerebellar pathways, but malfunctioning of other systems can in a lesser extent also be responsible for this phenotype.[2] For example, sensory (proprioceptive) dysfunction resulting from sensory peripheral neuropathies, dorsal root ganglionopathies or posterior spinal column involvement can produce an ataxic syndrome.[38] Although the careful examiner can often find clinical clues to distinguish cerebellar ataxia from sensory ataxia, detection of the latter is usually difficult. Other possible causes (or at least mimics) of ataxia are vestibular dysfunction and lesions of the prefrontal cortex.[2]

Ataxia is a symptom and concerns no final diagnosis. It can be caused by many neurological diseases, acquired as well as genetic. For both, the differential diagnosis is very broad and clinical presentation is often nonspecific which makes obtaining a correct diagnosis a clinical challenge.[39, 40] Despite thorough investigations, many patients with ataxia remain unexplained in the end.[41, 42] Causes of acquired ataxias are chronic alcohol use, toxic agents, immune-mediated inflammation, vitamin deficiency and chronic infection of the central nervous system.[43]

The hereditary ataxias are a heterogeneous group of rare degenerative disorders that can be inherited in an autosomal dominant, autosomal recessive, X-linked manner or through maternal inheritance if part of a mitochondrial genetic syndrome.[44] The largest group, the spinocerebellar ataxias, were traditionally classified according to disease phenotype, depending on first gene locus description and on inheritance mode.[5] The autosomal dominant spinocerebellar ataxias (SCAs) or autosomal dominant cerebellar ataxias (ADCAs) were divided in ADCA type 1, in which cerebellar ataxia is variably associated with other neurologic features, and ADCA type 2, in which there was relatively pure cerebellar ataxia.[45] For the recessive ataxias, two paralleled classifications exist, namely the autosomal recessive spinocerebellar ataxias (SCARs) classification and the autosomal

recessive cerebellar ataxia (ARCA) classification.[5] A third, smaller group are the somewhat arbitrarily categorized ‘spastic ataxias’ (SPAX), presenting with combined ataxia and spasticity.[5]
In these clinicogenetic disease classifications, the different subtypes of ataxia are often named ‘SCA’ or ‘SCAR’ with a number, the latter usually referring at the order in which they were first described (or when the disease locus or causative gene of the subtype was identified).[46] The disease name or type of ataxia (e.g. SCA1, SCA2) is essentially a more clinical description; with for each subtype certain more or less distinguishing features. For most subtypes, the corresponding genes or loci have now be identified. The disease name and gene are often used interchangeably.
However, several problems exists with these classification systems: (1) some loci are erroneously assigned, (2) some loci are duplicated, (3) some loci are missing and (4) some loci or genes are unconfirmed.[5] For example, some recessive ataxias are not in the SCAR and neither in the ARCA list (e.g. Friedreich’s ataxia or ataxia with oculomotor apraxia type I (AOA1)) or are listed in only one of them (e.g. ataxia with oculomotor apraxia type 2 (AOA2) is only in the SCAR classification). Moreover, as it turns out, some dominant ataxias can also be inherited in a recessive manner or the other way around (e.g. *AFG3L2)* and some ataxias can present as a pure of complicated form.[47] More importantly, as facilitated by the growing implementation of MPS, it is increasingly appreciated that overlap with other classifications such as HSP exist, as is discussed below. Therefore, the old linkage-inspired and phenotype-driven classification system seems to be outdated and a more modular, genotype-inspired system seems to be in order.[5]

The most common cause of hereditary ataxias are repeat expansions.[46] Repeat expansions are repeating units of three or more nucleotides that occur in tandem.[48] In the general population, wild-type alleles are polymorphic, meaning that a variable number of repeat units exist in the normal population. However, the number of repeats can increase when they are passed from generation to generation, undergoing expansion. This leads to an extent far beyond the normal polymorphic range, causing gene expression and function abnormalities.[48] A feature of many repeat expansion disorders is anticipation. This describes the way that a disease may get more severe going down the generations, for example with an earlier and earlier age of onset.[48, 49] The reason for this is that the expanded repeats are very unstable and may expand further on transmission. Since the size of the expansion is often correlated with severity and/or age at onset, this can lead to a more severe disease in future generations. Currently, around 40 repeat expansion disorders are known and interestingly, almost all of these conditions are primarily neurological (Table 1).[48, 49] Many of these are so-called polyglutamine disorders, carrying expanded repeats of the trinucleotide CAG which codes for the amino acid glutamine, but other configurations exist. An example is the recently identifed biallelic intronic AAGGG repeat expansion in the *RFC1* gene as cause of the cerebellar ataxia, neuropathy and vestibular areflexia syndrome (CANVAS).[50] Although this syndrome was

typically believed to consist of a triad of vestibular hypofunction, cerebellar dysfunction and sensory neuropathy, it was now shown that not all patients necessarily display the full spectrum of the disease. Sensory neuropathy was found to be a common feature, but further evidence is needed to further characterize the exact phenotype of these patients.[51] This shows the importance of a correct genotype-phenotype correlation, since it is very likely that the disease is underdiagnosed if the full phenotypic spectrum of the disease is not appreciated.

Next to repeat expansion disorders, other subtypes are caused by other mutation types e.g. missense mutations, small deletions or insertions and nonsense mutations.[52] Globally, SCA3 is the most common autosomal dominant and Friedreich the most common autosomal recessive ataxia, although population differences exist (Fig. 4).[46] Ataxia genes are involved in various molecular pathways, such as but not exclusively nuclear functioning (e.g. transcription regulation and DNA repair), protein folding, synaptic transmission, lipid biosynthesis and mitochondrial functions.[53, 54]

Unfortunately no cure exists for hereditary ataxias and management is mainly symptomatic.[52] However, gene therapies for some subtypes are currently under investigation.[55]

Disease	Omim No.	Mode of inheritance	Location of gene	Location of repeat	Repeat sequence
SCA1	164400	AD	6p23	Exon 8	$(CAG)_n$
SCA2	183090	AD	12q24	Exon 1	$(CAG)_n$
SCA3	109150	AD	14q24-q31	Exon 10	$(CAG)_n$
SCA6	183086	AD	19p13	Exon 49	$(CAG)_n$
SCA7	164500	AD	3p21-p12	Exon 3	$(CAG)_n$
SCA8	603680	AD	13q21	UTR	$(CTG)_n$
SCA10	603516	AD	22q13	Intron 9	$(ATTCT)_n$
SCA12	604326	AD	5q13	Promotor	$(CAG)_n$
SCA17	607136	AD	6q27	Exon 3	$(CAG)_n$
DRPLA	125370	AD	12p13	Exon 5	$(CAG)_n$
SCA36	614154	AD	20p13	Intron 1	$(GGCCTG)_n$
SCA37	603448	AD	1p32.2-p32.1	5' UTR	$(ATTTC)_n$
Friedreich ataxia	229300	AR	9q13-q21.1	Intron 1	$(GAA)_n$
Fragile X syndrome/FXTAS	309550	XL	Xq27.3	5'UTR	$(CGG)_n$
CANVAS	102579	AR	4p14	Intron 2	$(AAGGG)_n$
FTD/ALS	105550	AD	9p21	Intron 1	$(GGGGCC)_n$
SBMA	313200	XL	Xq11	Exon 1	$(CAG)_n$
Huntington disease	143100	AD	4p16	Exon 1	$(CAG)_n$
Myotonic dystrophy 1	160900	AD	19q13	3'UTR	$(CTG)_n$
Myotonic dystrophy 2	602668	AD	3q21	Intron 1	$(CCTG)_n$
Progressive myoclonic epilepsy	254800	AR	21q22.3	Promotor	$(CCCCGCCCCGCG)_n$

Table 1: diseases associated with pathogenic expanded nucleotide repeats.
This is not an exhausitive list. The first 15 diseases have ataxia as one of the primary symptoms, the others are examples of other neurodegenerative disorders. Adapted from Read et al.[49]
Abbreviations: AD, autosomal dominant; ALS, amyotrophic lateral sclerosis AR, autosomal recessive; CANVAS, cerebellar ataxia, neuropathy and vestibular areflexia syndrome; DRPLA, dentatorubral pallidoluysian atrophy; FTD, frontotemporal dementia; SBMA, spinobulbar muscular atrophy; SCA, spinocerebellar ataxia, UTR, untranslated region; XL, X-linked.

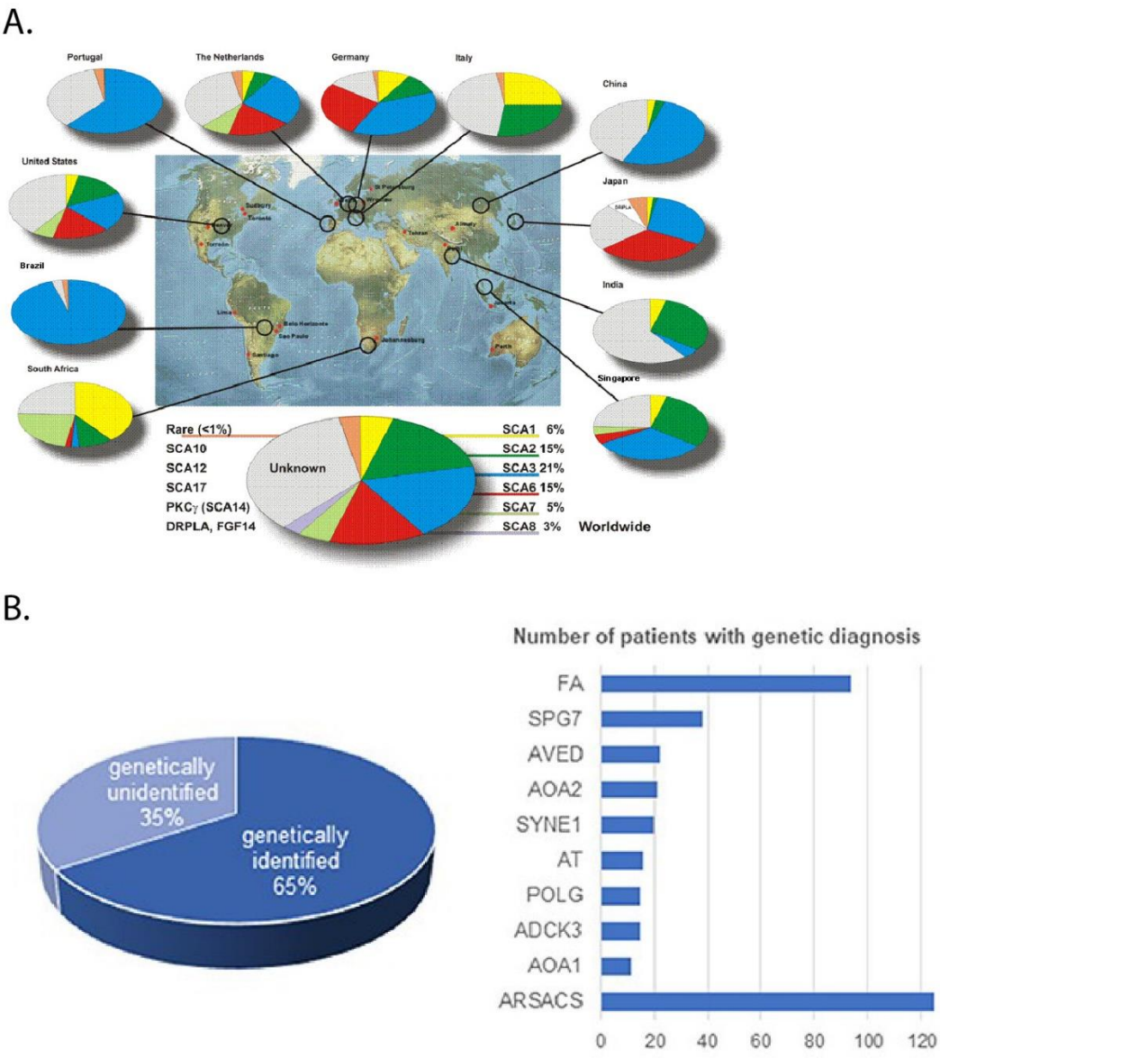


Figure 4: Frequencies of the most frequent causative genes in autosomal dominant cerebellar (A) and autosomal recessive (B) ataxia.

(A) Worldwide distribution of SCA subtypes. Figure reproduced from http://www.genereviews.org; copyright © 1993-2021 University of Washington, with permission.

(B) Genetic diagnoses in the ARCA registry, a multicenter registry for ARCA with contributing centers from (mainly) Europe as well as the United States, Canada, Iran, Tunis, Australia and New-Zealand. Currently >800 patients have been included. The pie plot indicates how many patients received a genetic diagnoses, the bar plot indicates the list of the 10 most frequent genetic causes (absolute numbers**). Except for the enrichment of ARSACS—which is an overrepresentation due to the major contribution of participating sites in Quebec**—the ARCA Registry provides prevalence data that are generally consistent with expectations from the literature.

Reproduced from Traschütz A, Reich S, Adarmes AD et al; available under Public License.[56] © 2021 Traschütz, Reich, Adarmes, Anheim, Ashrafi, Baets, Basak, Bertini, Brais, Gagnon, Gburek-Augustat, Hanagasi, Heinzmann, Horvath, de Jonghe, Kamm, Klivenyi, Klopstock, Minnerop, Münchau, Renaud, Roxburgh, Santorelli, Schirinzi, Sival, Timmann, Vielhaber, Wallner, van de Warrenburg, Zanni, Zuchner, Klockgether, Schüle, Schöls, PREPARE Consortium and Synofzik.

Hereditary spastic paraplegia

Spastic paraplegia (SPG) is characterized by hyperreflexia, spasticity and often loss of muscle strength in the lower limbs. Similar to ataxia, SPG is a symptom and not a disease, and many acquired and genetic etiologies are known to date.[4] Examples of acquired causes include structural abnormalities of the spinal cord, infections of the central nervous system and vitamin deficiencies.[57] For the hereditary spastic paraplegias, all modes of inheritance (autosomal dominant, autosomal recessive, X-linked and more rarely mitochondrial inheritance) have been described.[58] As is the case for ALS, and also for ataxia, patients can present with or without a suggestive familial history. Sporadic cases can still exhibit a genetic etiology, e.g. because of autosomal recessive variants, de novo mutations or AD variants with reduced penetrance. However, the diagnostic rate is lower than in familial cases (Fig. 5) albeit the difference is less pronounced compared to ALS. A possible explanation is that some patients actually have an unknown acquired etiology and not a genetic one. A classification system based on phenotype and locus description exists, however with the same limitations as the ones described for ataxia above.[5] Also here, it concerns a clinicogenetic classification where disease names ('SPG' or 'HSP' and number) and the corresponding genes are often used interchangeably. More than 80 genetic types of HSP have been defined to date, with SPG4 (*SPAST*) being the most common subtype (Fig. 5).[59] Causative genes are involved in multiple molecular pathways, such as axonal transport, control of endoplasmic reticulum morphology, mitochondrial function, myelination, protein folding and degradation, vesicle transport, lipid metabolism and purine nucleotide metabolism.[57, 60] However, it is not known if these different mechanisms are part of one shared biochemical process for which (distal) corticospinal axons are more vulnerable, or if they exist separately from each other. Treatment is supportive and symptomatic, for example medication to reduce muscle spasticity and physical therapy. Thus far, no gene-specific therapy is available, but studies in animal models and neurons derived from induced pluripotent stem cells have provided some potential therapeutic targets for specific subtypes of HSP.[61]

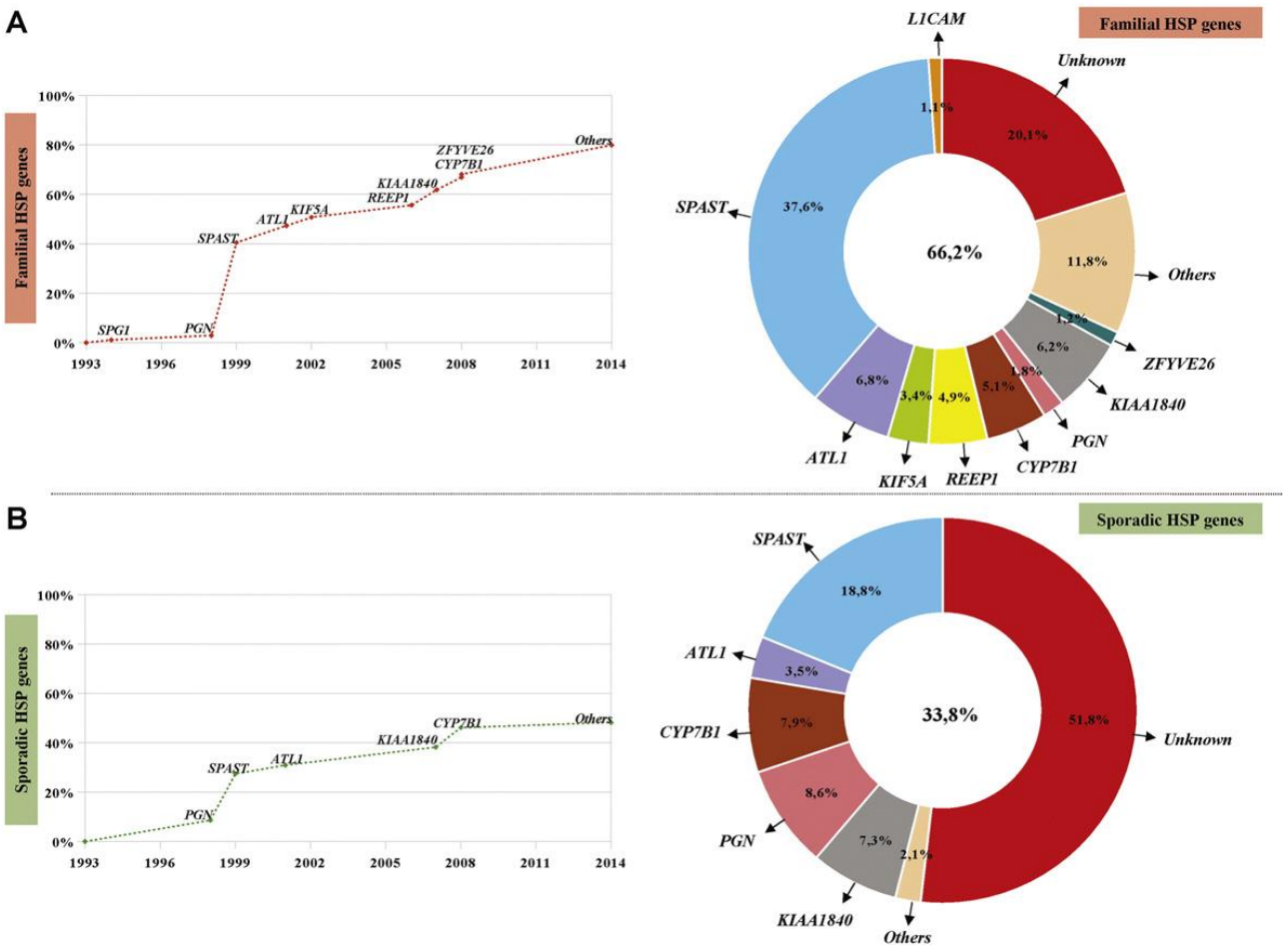


Figure 5: Timeline of gene discoveries and frequencies of the most frequent causative genes in familial (A) and sporadic (B) HSP.
Epidemiological data of the HSP Italian network (total of 1707 affected subjects from all over the world).
The number in the circle indicates the proportion of patients that are familial (top) and sporadic (bottom).
Reproduced from Lo Giudice et al. (2014); with permission.[60]

Heterogeneity and overlap in disorders affecting the motor system

Phenotypic heterogeneity and overlap

Although it may seem from what is written above as if each of the described disorders unilaterally affect a distinct part of the motor system, the truth is less straightforward (Fig. 1). While 'pure' forms of hereditary ataxia and HSP exist, affecting almost uniquely the cerebellar or upper motor neuron system respectively, many patients are known to exhibit so called 'complex' phenotypes. These include additional involvement of other parts of the motor system, such as the peripheral nervous system (including lower motor neurons), the extrapyramidal system, ataxia in HSP and upper motor neuron involvement in ataxia. Also other neurological systems such as cognition can be affected and other signs such as hearing loss, retinitis pigmentosa, optic atrophy or radiological abnormalities (e.g. on MRI brain) can be present.[52, 58] Overlap between the phenotypes exists, especially for ataxia and HSP, with as ultimate example the 'spastic ataxias', where patients exhibit both profound ataxia and spasticity.[5] Also for CANVAS and *RFC1* repeat expansions, typically associated with

ataxia, it is now increasingly appreciated that motor neurons are involved as well (both higher and lower motor neurons).[62] In and among the different subtypes of hereditary ataxia and HSP themselves, overlap and heterogeneity dominate the clinical picture.[4, 5, 52, 58] What exactly is causing this phenotypic heterogeneity, for example secondary genetic factors and/or environmental factors, is not yet fully understood. Even patients with mutations in the same gene, or with the exact same mutation, can show different phenotypes with involvement of different systems.[51, 63] Moreover, disease severity and age at onset can show striking inter- and intrafamilial variability. Lack of discriminating clinical or radiological features pointing to specific subtypes make it often impossible to predict the underlying molecular diagnosis based on the phenotype of the patient.

Compared to hereditary ataxia and HSP, ALS is a more 'pure' disorder. However, it already affects two different parts of the motor system in its most traditional form, namely the lower motor neurons and the upper motor neurons, the latter in overlap with HSP. Furthermore, a rare form of motor neuron disease (so called 'finger extension weakness and downbeat nystagmus motor neuron disease syndrome' (FEWDON)) with cerebellar symptoms exists. Downbeat nystagmus, oculomotor findings such as skew deviation, saccadic dysmetria and abnormal smooth pursuit has been described in FEWDON.[64] Additionally, as stated above, up to 50% of ALS patients develop cognitive and/or behavioral dysfunction as part of a spectrum with FTD.

Genetic heterogeneity and overlap

Although already appreciated before, the advent of MPS has uncovered a completely new level of genetic heterogeneity and overlap in and between disorders affecting the motor system. MPS is an unbiased approach, free from prior clinical conceptions, facilitating the discovery of both novel and overlapping genes associated with these diseases.[5, 65] For example, *SPG7* mutations were traditionally associated with SPG but are now known to cause ataxia and mixed spastic-ataxic phenotypes as well. [66, 67] Another example is *KIF5A*, known to cause autosomal dominant SPG (SPG10), but now also detected in patients with marked spastic ataxia.[68] *ATXN2* repeat expansions are a long known cause of hereditary ataxia, but intermediate repeat expansions have been identified in ALS as well and the same is true for *ATXN1*.[25, 69] The list can further be complemented with, but is not limited to, *KIF1C* (ataxia and HSP), *PNPLA6* (ataxia and HSP), *GBA2* (ataxia and HSP) and *SETX* (ataxia, ALS and distal hereditary motor neuropathy with upper motor neuron signs).[4, 70, 71]

Not only on gene-level itself, but also concerning the underlying pathophysiological mechanisms, evidence is arising that shared pathways and mechanisms underlie these diseases. For ataxia and HSP, this could already be deduced from the examples of involved pathways given above.[5] Also

in ALS, some of the same pathways such as axon transport, mitochondrial function and DNA repair seem to play a crucial role in the pathogenesis.[18]

Vulnerability of the motor system

As argued for above, there is extensive phenotypic, genetic and mechanistic overlap (which are of course not three separate entities but are instead all intertwined) between the disorders affecting the motor system. This arises suspicion for a shared vulnerability of the motor system, which places the involved phenotypes along a continuum. Especially for ataxia and HSP, it is nowadays acknowledged that they are actually part of a spectrum where both the corticospinal tract axons and the cerebellar circuits are vulnerable to disturbances in the same molecular pathways.[5] Their phenotypic expression is somewhere along this spectrum, being pure SPG, pure ataxia or mixed forms of spasticity and ataxia. This raises questions about the classification systems of these disorders. Given the genetic and phenotypic overlap, where it is sometimes even difficult to pick the 'most dominant system involved', the classic classification systems with categorical distinction between autosomal dominant ataxia, autosomal recessive ataxia and HSP seem artificial.[5, 72] Classification systems mainly based on the main phenotypes are outdated, and it seems right to at least involve the genotype more prominently, although this is difficult for patients for whom no molecular diagnosis can be established. Arguments are made for other classification systems, for example with 'modular phenotyping', an unbiased approach that allows a nuanced phenotypic expression along the spectrum of ataxia and spasticity, or classifications involving the phenotype, inheritance mode and gene alteration.[5, 72]

For ALS, atypical families with long disease courses have been described, for which one might wonder if these patients have ALS, HSP or something between the two.[73] Similarly, patients with FEWDON have an unusual slow disease progression, more typically for hereditary ataxia than for ALS.[64] Also here, a certain spectrum seems to exist, although less obvious than is the case for ataxia and HSP.

Generally, given the strong genetic basis for all these disorders, it seems plausible that there are certain genetic determinants of motor neuron vulnerability. The progress in molecular genetics opens opportunities to dissect this shared genetic architecture, which is an indispensable step for developing novel therapeutic strategies for these severe incurable disorders.

Next generation sequencing

A revolutionary approach

Next generation (NGS) or massive parallel sequencing has revolutionized genomic research. In contrast to 'first generation' techniques as Sanger sequencing, MPS can sequence millions of small DNA fragments in parallel, making it a fast and cost-efficient way to analyze large amounts of DNA.[74]

Generally, three main approaches can be distinguished: (a) targeted sequencing (TS), a method where only certain genes or regions of the genome of interest are sequenced; (b) whole exome sequencing (WES), where all coding regions from the human genome are analyzed, which form around 1-2% of the total genome, and (c) whole genome sequencing (WGS), where the total genome, including coding and non-coding regions of the genome is sequenced.[75] For all three approaches, the first step includes library preparation (Fig. 6A). The DNA of interest is broken down in numerous smaller fragments, using physical methods such as acoustic shearing and sonification, enzymatic methods or chemical methods.[76] Next steps include end-repair, A-tailing of the 3' ends (to facilitate ligation to sequencing adapters) and ligation of adapters. When samples are multiplexed, a unique index is embedded in the DNA construct through ligation of indexed adapters. Alternatively, indexing is performed through PCR amplification using indexed primers later on. After library preparation, an extra step is implemented when only a part of the genome is sequenced, e.g. TS or WES. Because not the whole genome is sequenced in these approaches, the DNA sample must be enriched for the DNA regions of interest; a process that is called 'capturing'. This is accomplished by hybridization of target (exon) sequences to biotinylated DNA or RNA baits, followed by washing away the sequences that are not hybridized.[77] The next step is amplification of the DNA, for which different techniques can be performed.[74] One of them is so called 'bridge amplification'. In this setting, DNA oligonucleotides that are complementary to the adapters in the library are attached to a flow cell (Fig. 6B). When the DNA fragments of the library are added to the flow cell, they hybridize with one of the oligonucleotides on the flow cell surface. After replication by DNA polymerase, the double-stranded fragment is denatured and the original strand is washed away. Because adapters were ligated to both ends of the DNA fragments, the one on the free end of the strand can now bind on another oligonucleotide of the flow cell, forming a bridge. Again DNA polymerase replicates this fragment forming a double-stranded DNA molecule, which is still bridged. After denaturation however, two single stranded copies of the DNA exist, each anchored on one end to the flow cell. This process is repeated several times, resulting in clusters of identical DNA molecules on the flow cell. When the desired amount of DNA has been reached, the actual sequencing can be performed (Fig. 6C). Again, the exact technique can differ, but here we shortly will discuss the four-colour cyclic reversible termination method that is commonly used.[74] In the

first step, a DNA polymerase adds just one fluorescently modified nucleotide. Since all DNA molecules in one cluster are clones, the same nucleotide is added in each individual cluster. After washing away the unincorporated nucleotides, imaging is performed to identify the incorporated nucleotide per cluster. Then the terminating group (preventing further DNA synthesis after incorporation of the fluorescent nucleotide) and the fluorescent dye are cleaved, so the process can restart and the next fluorescently modified nucleotide can be incorporated. Finally, when sequencing of the DNA fragments has been performed, bio-informatic analysis can begin (Fig. 6D). After quality control, the reads (this is the sequence of the DNA fragments) are aligned to a known reference sequence. Next, variants are called, which means that differences between the reads and the reference genome are identified, annotated and filtered e.g. against variants known to exist frequently in the general population. The final step is the manual interpretation of the variants, where different lines of evidence are used to estimate if a variant is (rather) benign or pathogenic and relevant for the phenotype of the patient.

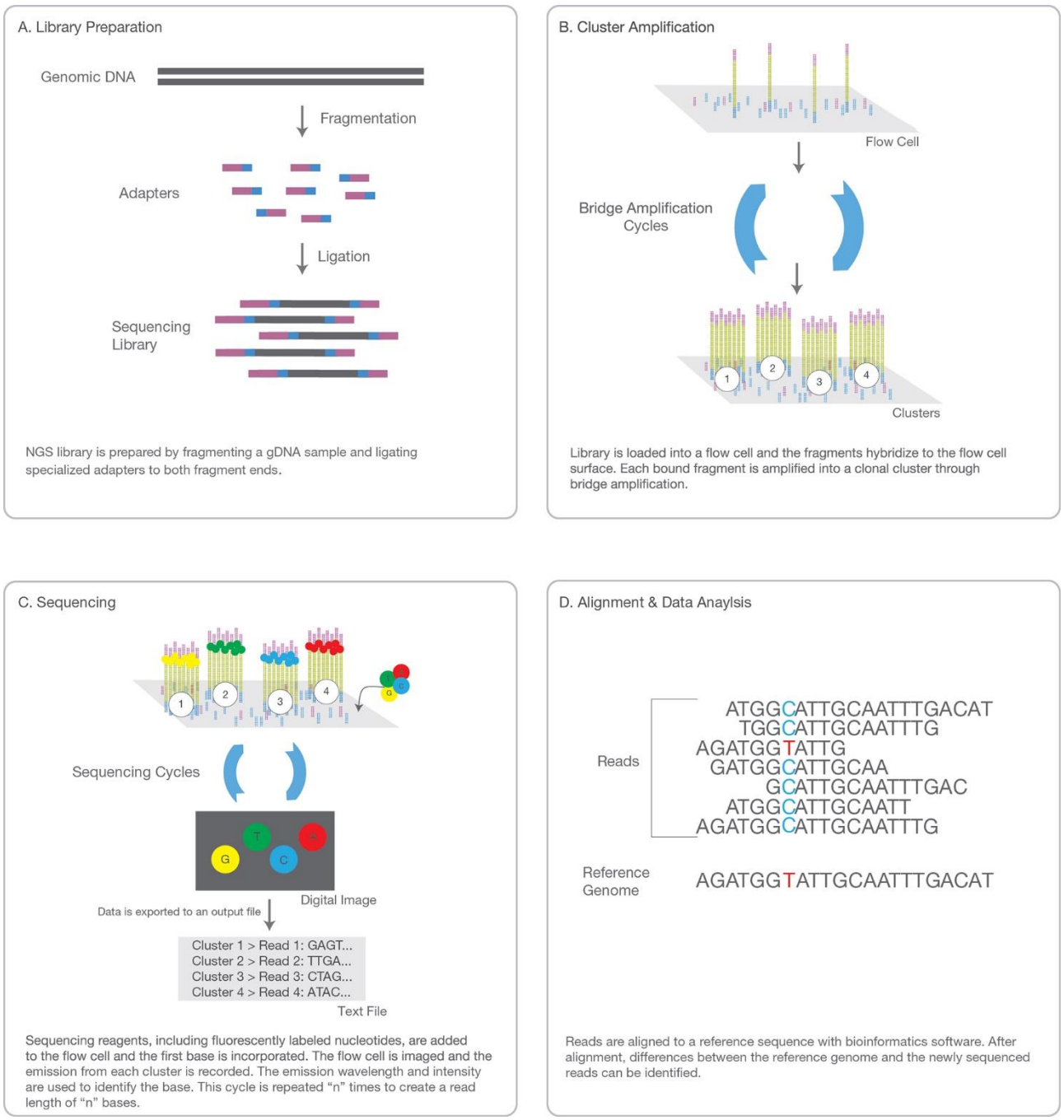


Figure 6: MPS workflow.
Courtesy of Illumina, Inc.

Advantages and limitations

The major advantage of MPSis the possibility to sequence large amounts of DNA in a fast and cost-efficient way. Where it took the Human Genome Project 13 years to completely sequence the human genome by massive application of Sanger sequencing, a whole genome can now be sequenced in one single run. This is especially an advancement for heterogeneous disorders with many known genes as ataxia and HSP, where gene-by-gene testing of all known genes is an almost impossible and very expensive journey to undertake. With MPS, all these genes can be sequenced and

analyzed with one single test. Another advantage is that MPS is unbiased from prior clinical and diagnostic preconceptions, making it possible to uncover new genotype-phenotype correlations that previously stayed under the radar because only patients with phenotypes typically known for the genotype were screened (e.g. ataxia in patients with *SPG7* mutations as explained before). Furthermore, searching for novel genes is possible if WES or WGS is performed in unexplained patients.

TS has been promoted because of its low costs, short turnaround time and low rate of incidental findings.[78] However, a major drawback is that it does not allow for post-hoc analysis of extra genomic regions when new genes are linked with the phenotype of (unexplained) patients. Since WES sequences the whole exome, it has the advantage that all known genes can be analyzed, even in retrospect when originally only a gene panel associated with the phenotype was evaluated. WGS evaluates the whole genome and thus also the intronic regions, unlike WES and TS. This is of course the most comprehensive way to evaluate a patient's DNA, but currently, the interpretation of intronic variants remains very challenging and the cost of WGS is still higher than for the other approaches. Although the read depth (this is the number of reads per nucleotide) is generally lower for WGS than for WES (than for TS), the read depth in WGS is more similar or stable throughout the genome.[79] This has the advantage of a more reliable detection of copy number variations.[78] Even for the detection of repeat expansions, an increasing number of tools becomes available, especially for WGS data.[80] One of the most recent ones, ExpansionHunter Denovo, is able to detect repeat expansions without prior knowledge of the genomic coordinates and even if there are nucleotide composition changes within the expanded repeats.[81] However, GC-rich regions, repetitive regions and detection of structural variation remain difficult for second generation sequencing techniques as discussed above. In the future, 'third generation sequencing' or 'long-read sequencing' will provide more possibilities for these matters.[82]

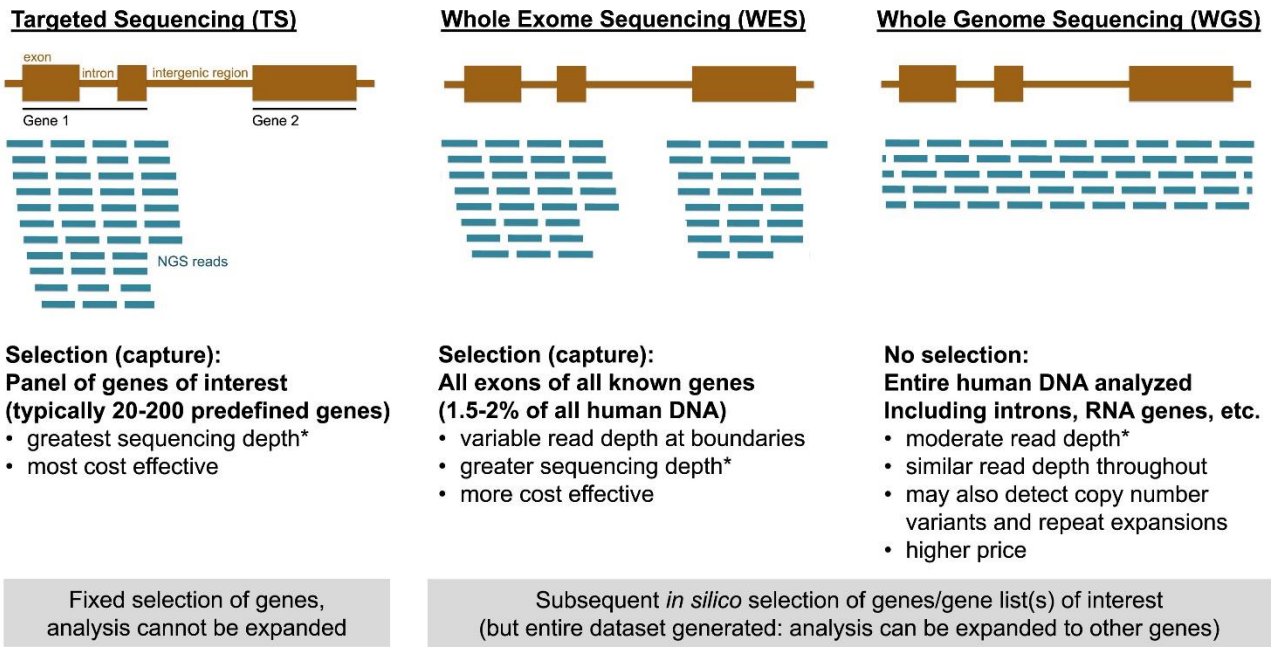


Figure 7: Overview of targeted sequencing, whole exome sequencing and whole genome sequencing.
Reproduced from Gorcenco et al. (2020); with permission.[79]

MPS in disorders affecting the motor system

For ALS, clinical genetic testing nowadays is still mainly performed as single gene testing for the four most commonly mutated genes (*C9orf72, SOD1, TARDBP* and *FUS*). It is common practice to offer gene testing for familial ALS, but much less so for sporadic ALS. In research however, MPS has gained in importance. Since it has been postulated that rare variants play an important role in the genetic architecture of ALS, the focus has shifted from GWAS-studies to MPS.[31] Several studies have been performed in ALS patients, mainly in familial ALS.[83-91] However, the number of patients is often small. Given the uncertain association of many genes with ALS (as discussed above), it is difficult to assemble gene panels resulting in different sets of genes being analyzed in distinct studies. Another major problem is the lack of consensus criteria to assess pathogenicity in variants for complex genetic disorders. The widely used criteria of the American College of Medical Genetics and Genomics (ACMG) are developed to detect rare, highly penetrant variants in Mendelian disorders and thus less suited for complex genetic disorders, where it is expected that more common and less penetrant variants also play an important role.[92] Here again, heterogeneous methods to attribute pathogenicity to variants are used throughout literature, often mainly based on variant frequency in population databases and/or controls and predicted functional consequences of the variant. All this makes it difficult to compare the results of different studies. A common theme however is the excess of rare mutations in ALS patients, and the co-occurrence of two (or more) variants in the same patients, suggestive for the oligogenic hypothesis.

Given the prospect of gene therapy for ALS, often targeted to genetic subtypes, the interest for genetic testing in ALS is increasing. However, implementing MPS for ALS in a diagnostic setting is challenging given the complex genetic architecture and the difficulties mentioned above. Especially for sporadic ALS, large studies describing genetic variation are rare and it is not sure what we can expect to find in this population.[91, 93] However, a better insight is essential for the interpretation of variants in a diagnostic setting, for the counseling of patients and their families and for therapy development for ALS.

For ataxia and SPG, MPS studies have been performed with yields generally between 10-55%.[53, 54, 65, 94-107] Results are more straightforward than for ALS given that hereditary ataxia and HSP are considered Mendelian disorders, but even here heterogeneity exists in many aspects, e.g. gene panel composition, attribution of pathogenicity, genes excluded before offering MPS and inclusion/exclusion criteria. This again complicates comparison between studies. Furthermore, selection of certain subpopulations as adult vs. pediatric onset and familial vs. sporadic patients in many studies make it unsure what to expect of MPS analysis in a real-life, clinical setting. Generally, mutations in common genes as *SPG4* and *SPG7* are a constant finding throughout the studies, but for most other genes heterogeneity dominates the spectrum.
For all disorders, but especially for ataxia and *C9orf72* in ALS, it should be emphasized that repeat expansions are currently not yet reliably identified in MPS data, at least not for diagnostic purposes. Therefore it remains important to exclude these disorders separately.

Chapter 2: Objectives

The objective of this PhD project was to gain more insight in the genetic determinants of disorders affecting the motor system. Therefore, we first aimed to map pathogenic mutations within each phenotype. We used retrospective chart analysis to determine disease-causing variants identified by gene-by-gene testing. We collaborated with the center for human genetics from UZ Leuven and participated in the gene panel analysis for ataxia and SPG patients. For ALS, we analyzed WGS data from a large international cohort of sporadic ALS patients. Secondly, we aimed to establish genotype-phenotype correlations. Therefore, detailed clinical data was collected from all patients and core clinical features as well as genetic information were logged into a database. Thirdly, we used this information to establish the gene frequencies in Belgium for ataxia and SPG, which can be important for guiding clinical DNA diagnosis. Lastly, we wanted to use the generated data to look for genetic and phenotypic overlap between the different disorders affecting the motor system.

(1) Amyotrophic lateral sclerosis – chapter 3

For ALS, the focus was placed on the sporadic population. Large studies providing a comprehensive analysis of genetic variation in these patients are rare, leaving us in the dark about what to expect from genetic testing in this population. As we participate in Project MinE, a large sequencing consortium for ALS, we were able to analyze WGS data from a large population of sporadic ALS patients.[108] Our aim for this project was to

- provide an inventory of pathogenic variation distilled from WGS data in sporadic ALS patients
- evaluate combinations of variants, in line with the oligogenic hypothesis
- evaluate the use of modified ACMG-criteria to weigh variant pathogenicity
- determine for which genes and variants previously reported as associated with ALS we can provide more evidence for a true association
- explore genetic overlap with other disorders affecting the motor system by analyzing a broad gene panel and evaluating the frequency of intermediate repeat expansions in *ATXN2* (typically associated with hereditary ataxia).

(2) Ataxia and SPG cohort – chapter 4

Although there is compelling evidence for the value of MPS in ataxia and SPG, studies describing the yield of MPS in an unselected, real-life population are rather scarce. Despite high yields in some studies, many patients remain unexplained. We retrospectively evaluated a cohort of ataxia and SPG patients presenting in the University Hospital in Leuven with no exclusion criteria based on family history or age of onset, reflecting a real-life clinical population. In this project we wanted to

- determine the yield of gene-by-gene testing
- determine the yield of MPS through gene panel analysis
- report genotype-phenotype correlations and novel variants; thereby also exploring genetic and phenotypic overlap between these disorders
- explore non-Mendelian inheritance by evaluating heterozygous variants in autosomal recessive genes associated with ataxia and SPG.

(3) Evaluate specific genotype-phenotype correlations – chapter 5 and 6

Although MPS offers major advances in clinically and genetically heterogeneous disorders as ataxia and SPG, establishing genotype-phenotype correlations remains of great value e.g. for the assembly of gene panels and the correct interpretation of the results. We focused on these correlations for two specific genes, *RFC1* and *RNF170*.

For *RFC1* repeat expansions, recently identified as the cause of CANVAS, our goal was to

- determine the yield of testing *RFC1* repeat expansions in a cohort of unexplained adult-onset ataxia cases
- provide genotype-phenotype correlations
- identify a subset of patients for whom genetic testing provides higher yields
- search for other repeat expansion constitutions.

For *RNF170*, we wanted to

- refine the genotype-phenotype correlation
- retrospectively analyse an in-house cohort for other (pathogenic) variants in the gene.

Chapter 3: Genetic variability in known ALS genes in sporadic ALS patients

Manuscript in preparation.

Abstract

BACKGROUND: With the advent of gene therapies for amyotrophic lateral sclerosis, there is a surge in gene testing for ALS. Although there is ample experience with gene testing for *C9orf72*, *SOD1*, *FUS* and *TARDBP* in familial ALS, large studies exploring genetic variation in all ALS-associated genes in sporadic ALS (sALS) are still scarce. Gene testing in a diagnostic setting is challenging given the complex genetic architecture of sALS with genetic variants with large and small effect sizes. Guidelines for interpretation of genetic variants in gene panels and for counseling of patients are lacking.

OBJECTIVE: To provide a thorough characterization of genetic variability in ALS genes by applying the ACMG criteria on whole genome sequencing (WGS) data from a large cohort of 6013 sporadic ALS patients and 2411 age- and sex- matched controls from Project MinE.

METHODS: We studied genetic variation in 90 ALS-associated genes and applied customized ACMG-criteria to identify pathogenic (P) and likely pathogenic (LP) variants. Variants of unknown significance were collected as well. In addition, we determined the length of repeat expansions in *C9orf72*, *ATXN1*, *ATXN2* and *NIPA1* using the ExpansionHunter tool.

RESULTS: We found *C9orf72* repeat expansions in 5.21% of sALS patients. In 50 ALS-associated genes, we did not identify any P or LP variants. In 5.89% of sALS patients, a P or LP variant was found (*C9orf72* repeat expansions not included). The most commonly mutated genes were *SOD1*, *TARDBP, FUS*, *NEK1*, *OPTN* and *TBK1.* Significantly more cases carried at least one P or LP variant compared to controls (OR 1.75; p-value 1.64×10^{-5}). Isolated risk factors in *ATXN1*, *ATXN2*, *NIPA1* and/or *UNC13A* were detected in 17.33% of cases. In 71.83% of patients, we did not find any genetic clues. A combination of variants was found in 2.88% in cases compared to 1.49% in controls.

CONCLUSION: This study provides an inventory of P and LP genetic variation in a large cohort of sALS. Overall, we identified P and LP variants in 5.92% of ALS patients in 38 known ALS genes. We found a significant enrichment of pathogenic variants in cases, although not for all ALS-associated genes evidence for association could be provided. In line with the oligogenic hypothesis, we found significantly more combinations of variants in cases compared to controls. Many variants of unknown significance may contribute to ALS risk, but diagnostic algorithms to reliably identify them are lacking. This work can serve as a valuable resource for counseling and for the assembly of gene

panels for ALS. Further characterization of the genetic architecture of sALS is necessary given the growing interest in gene testing in ALS.

Introduction

Amyotrophic lateral sclerosis (ALS) is a motor neuron disorder characterized by upper and lower motor neuron degeneration, which leads to progressive muscle weakness and wasting.[1] Up to 50% of patients develop extramotor symptoms, such as cognitive and/or behavioral dysfunction, as seen in frontotemporal dementia (FTD). [1] The disease is relentlessly progressive and most patients die between 2-5 years after disease onset but effective treatments are lacking.[1,2] ALS has a strong genetic component. In 5-10%, there is a familial history of ALS ('fALS'). Highly penetrant causal variants are found in ~70% of fALS patients, most commonly in *C9orf72*, *SOD1*, *TARDPB* and *FUS* which are responsible for about 40%, 20%, 4% and 3% of familial cases respectively.[3] The remaining 90-95% of patients present with apparently sporadic ALS (sALS) [4], but mutations in the same genes are found at lower frequencies. [5] Twin studies suggest a heritability in sALS patients of around 60%. [6] Both rare variants with a variable effect size, common variants with small effect size and combinations of such variants are thought to confer genetic risk in sALS patients. [7,8] Nevertheless, much of the genetic architecture of ALS remains unknown. During the last years, many efforts have been made to unravel the missing heritability. Genetic research has linked a considerable number of genes and variants to ALS through various different techniques. [7,9] However, the evidence for association is variable and some findings failed to be replicated in subsequent studies. [10] Furthermore, the significance of the distinct variants is often unclear (e.g. monogenetic with high penetrance, modifier, risk factor or in linkage with causal variant), especially in sALS patients. As the lifetime risk of ALS is about 1/400 and many of the disease-associated variants have an incomplete penetrance, such variants will also be found in control populations.[11] One of the difficulties in the interpretation of variants in complex genetic diseases like ALS is how to weigh the pathogenicity of variants, since consensus criteria are not available and different methods for scoring genetic variants are used. [12,13] With the advent of precision medicine for genetic subtypes of ALS, the interest in gene testing is increasing. It is common practice to offer gene testing for fALS, but much less so for sALS. Large studies systematically describing the genetic variation in sALS are rare.[14,15] However, a better insight in the level of pathogenicity of single variants is essential for the interpretation of variants in a diagnostic setting and for the counseling of patients and their families.

We analyzed whole genome data of a large cohort of 6013 sporadic ALS patients and focused on a set of 90 genes reported to be associated with ALS.[16] We implemented an automatic pipeline for classification of variants using modified criteria of the American College of Medical Genetics and Genomics (ACMG) to classify the variants.[17] We give a comprehensive overview of the pathogenic

(P) and likely pathogenic (LP) variants encountered in these ALS-associated genes, simultaneously validating the association of some of these genes with ALS in our cohort. In light of the oligogenic inheritance model of ALS, we also searched for the occurrence of combinations of such variants.

Methods

Patient selection

This study used whole genome sequencing (WGS) data of a total of 9600 samples of ALS cases and age- and sex- matched controls from Project MinE. Samples were mostly of people of European descent and collected from 17 centers across 13 countries: Belgium, The Netherlands, Switzerland, Spain, France, United Kingdom, Ireland, Israel, Italy, Portugal, Sweden, Turkey, and the United States. The diagnosis of ALS was made based on the El-Escorial criteria in the including center, which also provided clinical information. Case/control ratio was 2 to 1. For more detailed information on patient and control selection we refer to the pilot study of Project MinE.[16] A set of 8940 genomes passed QC, of which 6529 unrelated ALS index patients and 2411 unrelated healthy controls. fALS patients, defined as having a family member with ALS in the first or second degree, were excluded for this analysis (n=301). Patients with primary lateral sclerosis (n=74), progressive muscular atrophy (n=90) and progressive bulbar palsy (n=51) were also excluded. Our final cohort consisted therefore of 6013 sALS patients and 2411 controls.

Whole genome sequencing and bioinformatics analysis

DNA was extracted and sequenced on the Illumina Hiseq 2000 and Hiseq X platforms. Sequencing data was then aligned to hg37 using the iSAAC Aligner, and variants called using the iSAAC variant caller; both the aligner and caller are standard to Illumina's aligning and calling pipeline. Liftover to GRCh38 was performed using picard LiftoverVcf (v2.20.2-SNAPSHOT). QC was performed at sample and variant level as described previously. [18]

Gene selection

A list of 'ALS-associated genes' was selected from four databases (search conducted on April 28, 2021): Omim (search term 'amyotrophic lateral sclerosis'), the neuromuscular homepage (NMHP), the online version of the gene table and ALSod (genes with definitive and strong evidence).[19–22] All genes were included to provide an unbiased selection, regardless of the evidence that was available for an association with ALS. This resulted in a set of 91 genes (Supplementary Table 1). For one gene, *NOTCH2NLC*, the liftover failed, excluding this gene from further analysis. This brought the total number of eligible genes to 90. Additionally we randomly selected 90 genes with expression in brain as obtained from the GTEx Portal for comparison (Supplementary Table 2).[23]

The exact genomic regions corresponding to these genes were extracted from Ensembl (version 96) to which we added a buffer region of 2500 base pairs both upstream and downstream.[24] For each gene we selected the canonical transcript (Supplementary Table 1).

Variant extraction and annotation

We extracted all sequencing variants located in the genomic region of the selected genes from WGS data of patients and controls. Variants were annotated using ensembl-vep (v96.0-0). Vcfanno (v0.3.1) was used to add precalculated scores of the dbnsfp (v4.1a) and dbscsnv (v1.1) databases and add the allele frequency from Gnomad (v2.1.1 liftover_grch38).

Repeat expansions in *C9orf72* and *NIPA1* as well as intermediate repeat expansions in *ATXN1* and *ATXN2* were identified in our WGS data using the ExpansionHunter tool.[25] Cut-offs used were repeat lengths ≥ 30 for C9orf72, ≥39 for ATXN1, ≥33 for ATXN2 and ≥ for NIPA1. The SNP rs12608932 in *UNC13A* is a known risk factor for ALS.[26] Frequency of the CC genotype was distilled from our data.

Attribution of variant pathogenicity

To predict variant pathogenicity we used the software tool CharGer (Characterization of Germline variants), an open-source framework conducting a fully automated variant interpretation based on the ACMG-criteria, enriched with custom modules and a user-adjustable scoring system.[27] We partially customized the criteria and included some changes to make them more suitable for complex diseases. A summary of how we customized the ACMG criteria can be found in Supplementary Fig. 1 and 2 and Supplementary Table 1 and 3. After applying these modified criteria on the extracted sequencing variants, we selected all variants with a classification of pathogenic (class V) or likely pathogenic (class IV). Expansions in *ATXN1*, *ATXN2* and *NIPA1* and SNP rs12608932 in *UNC13A* were not considered pathogenic but were logged as risk factors.

Known mutations associated with ALS

We extracted variants associated with ALS in the 90 genes of interest from two databases: ClinVar and the NMHP.[20,28] We reclassified these variants as described in Supplementary Fig. 1 and selected the variants with a final classification as P and LP as ‘known ALS-associated mutations’. An overview of the number of variants reclassified per gene is provided in Supplementary table 4.

Statistical analysis

R (v3.6.3) was used to perform Fisher exact tests to compare variant frequency between cases and controls in our dataset assuming a dominant model. Odds ratios with zero cell counts in the control population were estimated by assuming the control population had a count of one, these ORs are denoted with a tilde. Multiple testing correction was performed using the holm method as implemented in p.adjust.

GnomAD

To calculate odds ratios compared to GnomAD v3, the variants that passed QC were downloaded from https://gnomad.broadinstitute.org and processed through our modified ACMG criteria, retaining only the pathogenic and likely pathogenic variants.[29] These were compared to the pathogenic and likely pathogenic variants present in our ALS cohort after removing the variants that were also removed in GnomAD. Fisher exact was then used as described above.

Graphical representation

Graphical representations called ‘lolliplots’ were created in R (version 3.6.0) using the Bioconductor package trackViewer (version 1.20.5).[27,30]

Ethical approval

The study was approved by the local ethical committee of the participating centers. All patients signed an informed consent. The Commission of medical ethics of UZ KU Leuven gave permission for sample collections (S50354) and for exome/genome sequencing (S52853).

Results

Our final cohort consisted of 6013 sALS patients and 2411 controls. The MinE project is set up to discover new ALS genes, so patients with known causal mutations were underrepresented. The main characteristics of the cohort are summarized in Table 1.

	Cases (n=6013)	**Controls (n=2411)**
Sex		
Female	2399 *(39.90%)*	1140 *(47.28%)*
Male	3614 *(60.10%)*	1271 *(52.72%)*
Age at onset		
Number of patients for whom data was available	5878	/
Mean age at onset (years)	60.35	
Age at time of bloodsample		
Number of patients for whom data was available	4883	2197
Mean age at time of bloodsample (years)	62.20	61.30
Site of onset		
Bulbar	1617 *(26.89%)*	/
Spinal	3916 *(65.13%)*	/
Generalized	201 *(3.34%)*	/
Thoracic/respiratory	107*(1.78%)*	/
FTD	7 *(0.12%)*	/
Unknown	165 *(2.74%)*	/
El Escorial (most recent value)		
Definite	1385 *(23.03%)*	/
Probable	2309 *(38.40%)*	/
Possible	515 *(8.56%)*	/
Suspected	57*(0.95%)*	/
Unknown	1747 *(29.05%)*	/
C9orf72* status		
Expanded	313 *(5.21%)*	7 *(0.29%)*
Intermediate	18 *(0.30%)*	4 *(0.17%)*
Normal	5659 *(94.11%)*	2397 *(99.42%)*
Inconsistent	22 *(0.37%)*	0 *(0%)*
Unknown	1 *(0.02%)*	3 *(0.12%)*

Table 1: Summary of the main characteristics of the MinE cohort.
The El Escorial criteria are diagnostic criteria for ALS based on clinical and electrophysiological evidence of lower motor and higher motor neuron degeneration.[31] Depending on the degree of lower and/or higher motor neuron disease, a patient can be classified as having 'definite', 'probable', 'possible' or 'suspected' ALS.
* C9orf72 status was determined by using the software tool 'ExpansionHunter'. 'ExpansionHunter'returns a range containing the probable number of repeats. Expanded means that the lower bound of this range is 30 repeats or more. Normal means that the upper bound of the range is less than 30 repeats. Intermediate means that the result is in between 'expanded' and 'normal'. Where available results of 'ExpansionHunter'were compared with wet lab results. 'Inconsistent' means that a discrepancy between the result of the 'ExpansionHunter'and wet lab results is found.

Abbreviations: FTD, frontotemporal dementia.

Using the ExpansionHunter tool, we identified 313 cases (5.21%) with a *C9orf72* repeat expansion. Additionally, we identified 365 P and LP sequencing variants in 354 (5.89%) cases and 87 P and LP sequencing variants in 83 controls (3.44%) (Table 2). Significantly more cases carried at least one P or LP sequencing variant compared to controls (OR 1.75; p-value $1.64x10^{-5}$).

When we exclude all patients and controls for whom rule PP5 (=reputable source reports variant as pathogenic) is triggered, we find 251 of 5910 cases with a variant, compared to 80 of 2408 controls (p=0,07). However, when in addition we exclude some less convincing genes (*NEFH*, *SYNE1*, *CACNA1H*, *SETX*), we find 194 of 5910 cases with a variant and 43 of 2408 controls with a variant ($p=3.8x10^{-4}$).

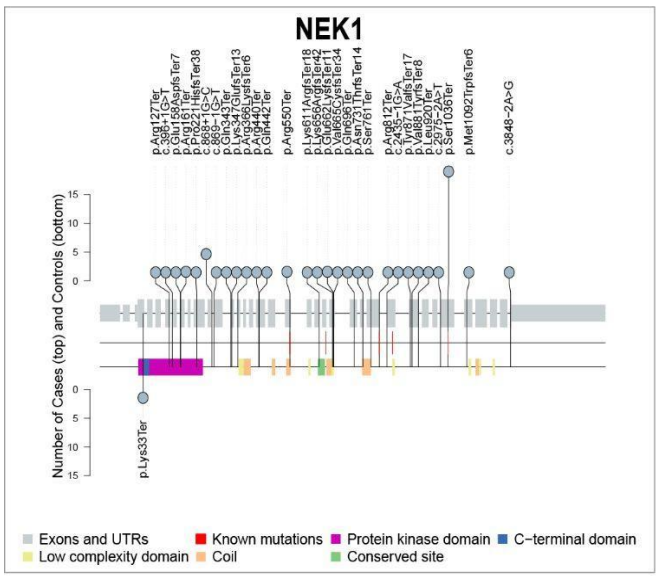
P and LP sequencing variants were observed in 40 of the 90 genes, that is in 38 genes in cases and in 24 genes in controls (Table 2, Fig.1, Supplementary Fig. 3). In 25 genes, a higher proportion of patients carried at least one P or LP variant compared to controls. For 10 genes the odds ratio (OR) was higher than 2. The results were significant for the genes *NEK1* (OR 21.84; p-value $2.57x10^{-7}$), *SOD1* (OR 7.66; p-value $3.15x10^{-4}$), *FUS* (OR ~6.43; p-value $9.40x10^{-3}$), *OPTN* (*OR* 4.83; p-value $1.58x10^{-2}$) and *TARDBP* (OR ~5.22; p-value $2.59x10^{-2}$). P-values remained significant only for *NEK1* and *SOD1* after correcting for multiple testing. However, by increasing the size of the control population by taking data from gnomAD, we could confirm these top 5 genes, obtaining even more significant p-values *(Table 2)*. Some other genes also reached significant p-values when compared to gnomAD after correction for multiple testing. However, only *TBK1* could be added to the list of significant genes using a more stringent genome wide correction for the ±20 000 genes in our genome. In 15 genes, a higher proportion of controls carried at least one P or LP variant. The p-value was significant only in *NEFH* (OR 0.30; p-value $3.26x10^{-2}$) but did not survive correction for multiple testing and was also not significant when controls from gnomAD were added. In 50 genes we did not find any P or LP variant.

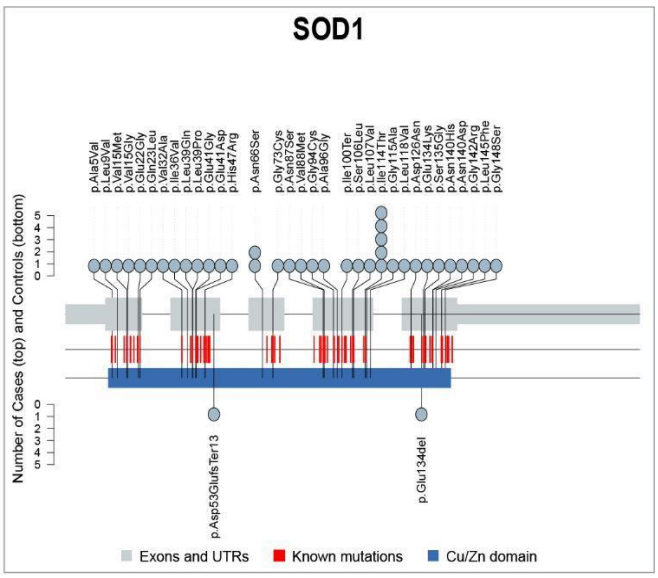
In the set of 90 randomly selected genes, we did not find significantly more pathogenic variants in cases or controls, neither in total nor on individual gene level (OR 1.46; p-value $3.39x10^{-1}$) (Supplementary table 5).

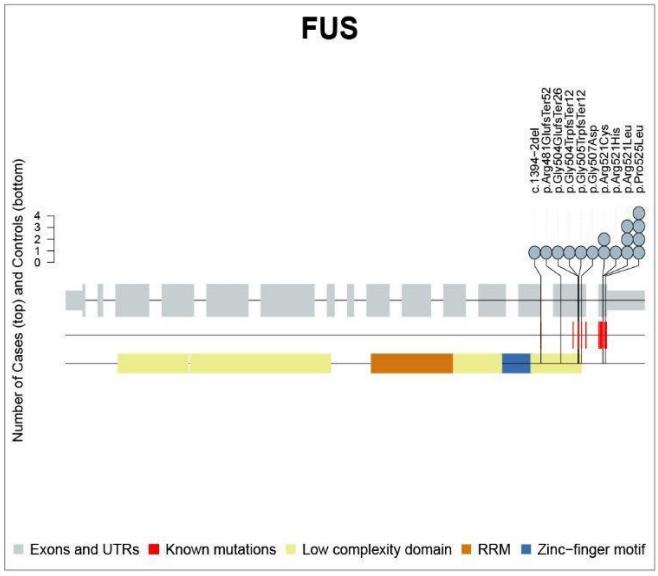
Of the variants associated with ALS in ClinVar and the NMHP, 367 were reclassified as likely pathogenic or pathogenic by applying our modified ACMG criteria. These variants were considered ‘known ALS-associated mutations’. In our cohort, 60 (23.0%) of the unique identified variants in cases were known ALS-associated mutations, the majority were novel variants.

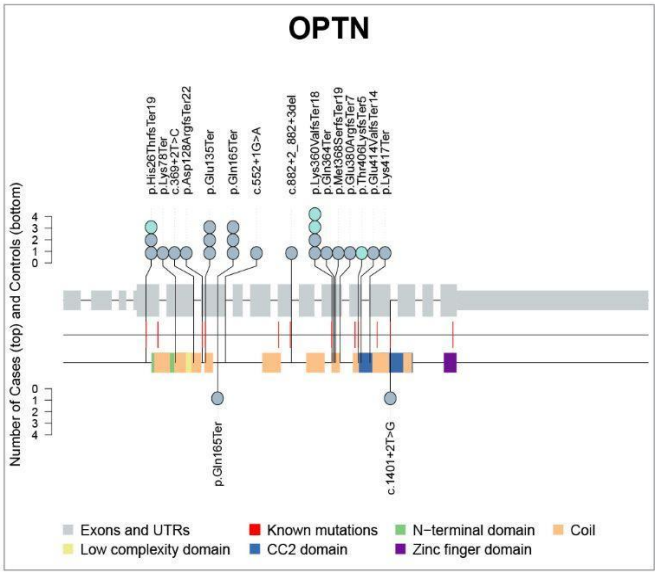
Taken together, 313 cases (5.21%) were explained by a *C9orf72* repeat expansion. A P or LP sequencing variant was observed in *SOD1*, *FUS, TARDBP, NEK1, OPTN* and *TBK1* in 38 (0.63%), 16 (0.27%), 13 (0.22%), 54 (0.90%) and 24 (0.40%) and 16 (0.27%) respectively. Of the remaining

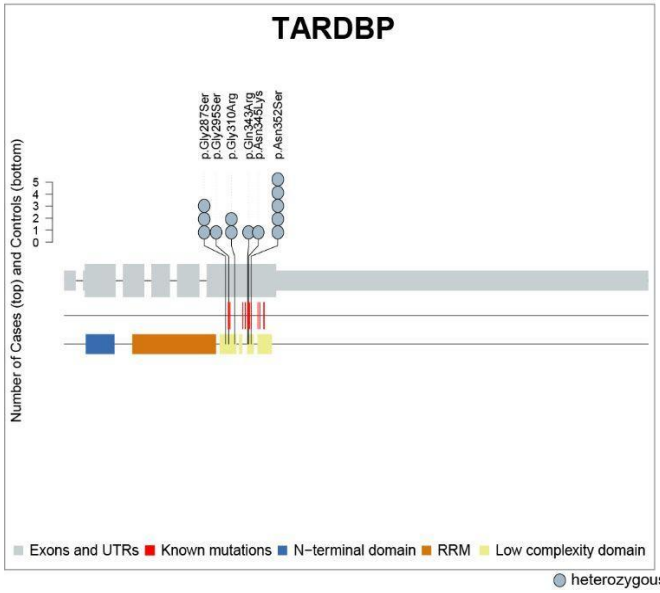
cases, 179 (2.98%) had a mutation in one of the other genes. In the remaining part of the cohort we detected one or more susceptibility factors in 1042 (17.33%) cases. In the other 4319 (71.83%) cases we did not find any genetic clue. In the control cohort, *C9orf72* repeat expansions were observed in 7 (0.29%) controls. P or LP sequencing variants were observed in *SOD1*, *NEK1, OPTN* and *TBK1* in 2 (0.08%), 1 (0.04%), 1 (0.04%) and 1 (0.04%) controls respectively. P or LP sequencing variants in the other genes were found in 76 (3.15%) controls (not statistically different from the proportion of 2.98% in cases (p=0.72). We found only susceptibility factors in 387 (16.05%) controls. No variants were found in 1934 (80.22%) controls.

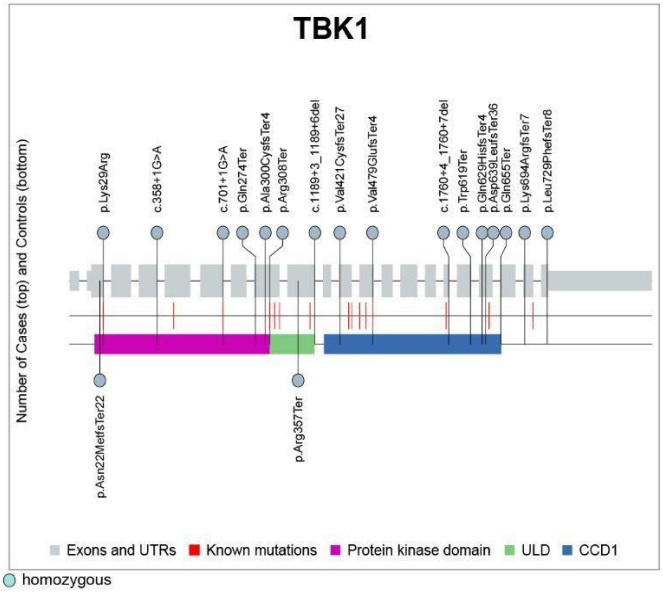












heterozygous homozygous

Figure 1: Lolliplots showing pathogenic and likely pathogenic variants encountered in NEK1, SOD1, FUS, OPTN, TARDBP and TBK1 in cases and controls.
Lolliplots show the mutational spectra of the displayed genes in cases (n=6013) and controls (n=2411).
For each gene, exons and UTRs are represented as grey bars, connected by introns shown as grey lines (reduced with a factor 1/100). Protein domains and motifs are shown as colored bars. Each lollipop represents a pathogenic or likely pathogenic variant encountered in cases (top) or controls (bottom). The absolute number of cases and controls with the variant is reflected by the height of the lollipop. Known ALS-associated mutations (from ClinVar and the NMHP) are shown as red lines.

Abbreviations: CCD1, coiled-coil domain 1; NMHP, neuromuscular homepage; RRM, RNA recognition motif domain ; ULD, ubiquitin-like domain; UTR, untranslated region.

To evaluate the incidence of combinations of variants, we took following variants into account: P and LP sequencing variants, *C9orf72* repeat expansions and ALS susceptibility factors including *ATXN1* and *ATXN2* intermediate repeat expansions, *NIPA1* repeat expansions and homozygosity for the C-allele of rs12608932 in *UNC13A*. A combination of two of these variants was found in 164 (2.73%) cases and in 34 (1.41%) controls (OR 1.96; p-value 2.31×10^{-4}) and a combination of three variants in 9 (0.15%) cases and in 2 (0.08%) controls (OR 1.81; p-value 0.74) (Fig. 2, Supplementary table 6).

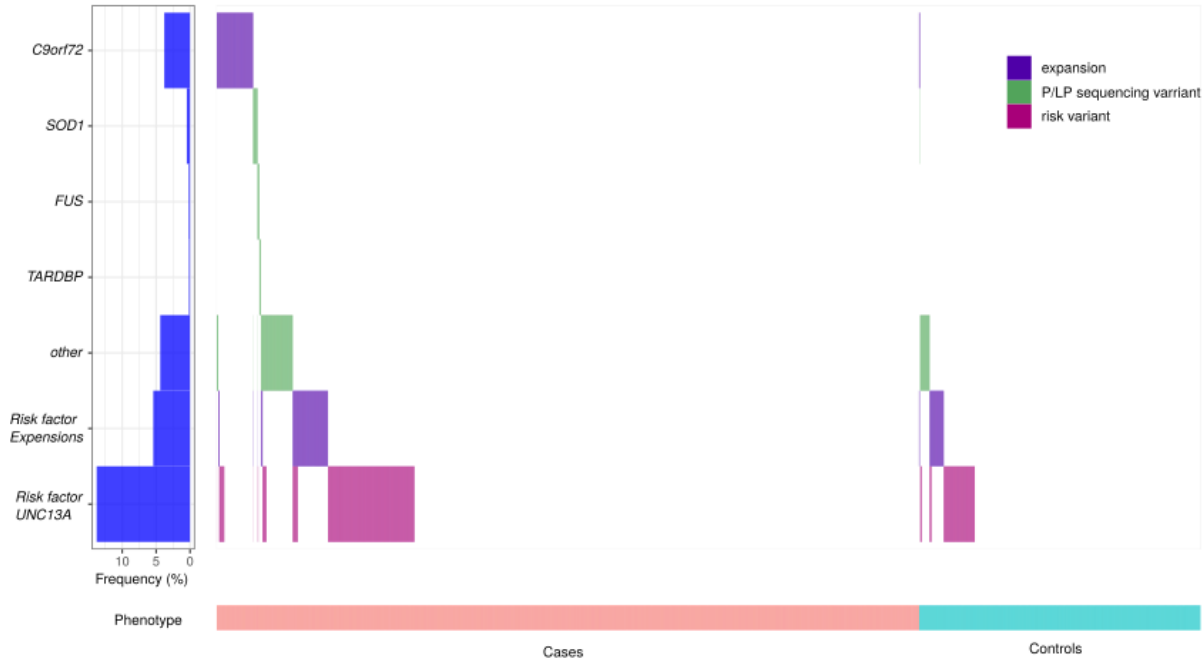


Figure 2: Waterfall plot: combinations of variants.
A waterfall plot shows for all cases (left) and controls (right) which variants they carry: repeat expansions in C9orf72; pathogenic or likely pathogenic sequencing variants in SOD1, FUS, TARDBP or in one of the other ALS-associated genes; risk factors such as (intermediate) repeat expansions in ATXN1, ATXN2 and NIPA1 or the CC-genotype of the SNP rs12608932 in UNC13A.Color code: in purple repeat expansions are shown, in green pathogenic (P) and likely pathogenic (LP) sequencing variants and in pink non-repeat expansion risk variants.

Abbreviations: LP, likely pathogenic; P, pathogenic.

							Gnomad				
Gene	**Case (%)**	**Control (%)**	**P-value**	**Adj p-value**	**Est. OR**	**95% CI**	**GnomAD (%)**	**P-value**	**Adj p-value**	**Est. OR**	**95% CI**
NEK1	54 (0.90)	1 (0.04)	$2.57x10^{-7}$	$1.03x10^{-5}$	21.84	3.75 - 875.45	52 (0.04)	$9.19x10^{-31}$	$3.58x10^{-29}$	11.84	7.93 - 17.68
SOD1	38 (0.63)	2 (0.08)	$3.15x10^{-4}$	$1.23x10^{-2}$	7.66	1.98 - 65.66	8 (0.01)	$4.47x10^{-34}$	$1.79x10^{-32}$	54.09	24.83 - 133.75
FUS	16 (0.27)	0 (0.00)	$9.40x10^{-3}$	0.36	~ 6.43	1.55 - Inf	5 (0.00)	$5.01x10^{-14}$	$1.85x10^{-12}$	36.19	12.67 - 126.77
OPTN	24 (0.40)	2 (0.08)	$1.58x10^{-2}$	0.58	4.83	1.2 - 42.2	20 (0.01)	$2.29x10^{-15}$	$8.72x10^{-14}$	13.62	7.2 - 26.01
TARDBP	13 (0.22)	0 (0.00)	$2.59x10^{-2}$	0.93	~ 5.22	1.22 - Inf	6 (0.00)	$1.12x10^{-10}$	$3.92x10^{-9}$	24.57	8.71 - 79.01
NEFH	6 (0.10)	8 (0.33)	$3.26x10^{-2}$	1.00	0.30	0.09 - 0.99	32 (0.02)	0.23	1.00	1.73	0.53 - 4.48
PRPH	9 (0.15)	0 (0.00)	$6.79x10^{-2}$	1.00	~ 3.61	0.79 - Inf	29 (0.02)	$2.85x10^{-3}$	$7.40x10^{-2}$	3.51	1.46 - 7.63
TBK1	16 (0.27)	2 (0.08)	0.12	1.00	3.21	0.75 - 28.84	3 (0.00)	$2.88x10^{-13}$	$1.04x10^{-11}$	52.92	14.76 - 285.4
ITPR2	12 (0.20)	1 (0.04)	0.13	1.00	4.82	0.71 - 205.93	47 (0.03)	$2.48x10^{-3}$	$6.70x10^{-2}$	2.89	1.4 - 5.54
SPG11	32 (0.53)	7 (0.29)	0.16	1.00	1.84	0.8 - 4.94	126 (0.09)	$1.28x10^{-6}$	$4.35x10^{-5}$	2.88	1.89 - 4.28
C19orf12	0 (0.00)	1 (0.04)	0.29	1.00	0.00	0 - 15.64	13 (0.01)	0.62	1.00	0.00	0 - 3.68
SQSTM1	0 (0.00)	1 (0.04)	0.29	1.00	0.00	0 - 15.64	10 (0.01)	1.00	1.00	0.00	0 - 5.05
CHRNA4	2 (0.03)	2 (0.08)	0.32	1.00	0.40	0.03 - 5.53	9 (0.01)	0.23	1.00	2.48	0.26 - 11.99
VCP	5 (0.08)	0 (0.00)	0.33	1.00	~ 2.01	0.37 - Inf	4 (0.00)	$3.36x10^{-4}$	$1.04x10^{-2}$	14.15	3.05 - 71.34
ALS2	7 (0.12)	5 (0.21)	0.34	1.00	0.56	0.15 - 2.24	32 (0.02)	$3.57x10^{-2}$	0.81	2.48	0.92 - 5.71
SYNE1	18 (0.30)	10 (0.41)	0.41	1.00	0.72	0.32 - 1.75	116 (0.09)	$3.72x10^{-2}$	0.81	1.76	1.01 - 2.9
CACNA1H	29 (0.48)	15 (0.62)	0.41	1.00	0.77	0.4 - 1.56	27 (0.02)	$1.91x10^{-3}$	$5.55x10^{-2}$	3.77	1.56 - 8.27

LRP10	1 (0.02)	1 (0.04)	0.49	1.00	0.40	0.01 - 31.48	13 (0.01)	1.00	1.00	0.87	0.02 - 5.8
TAF15	1 (0.02)	1 (0.04)	0.49	1.00	0.40	0.01 - 31.48	11 (0.01)	1.00	1.00	1.02	0.02 - 7
CHRNA3	9 (0.15)	5 (0.21)	0.56	1.00	0.72	0.22 - 2.74	58 (0.04)	0.12	1.00	1.76	0.76 - 3.57
BSCL2	3 (0.05)	0 (0.00)	0.56	1.00	~ 1.2	0.17 - Inf	5 (0.00)	2.20×10^{-2}	0.53	6.79	1.05 - 34.91
GLT8D1	3 (0.05)	0 (0.00)	0.56	1.00	~ 1.2	0.17 - Inf	18 (0.01)	0.24	1.00	1.89	0.36 - 6.46
UBQLN2	5 (0.08)	1 (0.04)	0.68	1.00	2.01	0.22 - 94.87	2 (0.00)	1.91×10^{-4}	6.10×10^{-3}	22.11	3.62 - 231.43
ATP13A2	5 (0.08)	3 (0.12)	0.70	1.00	0.67	0.13 - 4.31	45 (0.03)	0.60	1.00	1.26	0.39 - 3.16
ANXA11	8 (0.13)	2 (0.08)	0.73	1.00	1.60	0.32 - 15.52	20 (0.01)	1.33×10^{-3}	4.00×10^{-2}	4.53	1.72 - 10.74
SETX	11 (0.18)	5 (0.21)	0.79	1.00	0.88	0.28 - 3.24	39 (0.03)	1.92×10^{-3}	5.55×10^{-2}	3.19	1.47 - 6.36
ANG	2 (0.03)	0 (0.00)	1.00	1.00	~ 0.8	0.08 - Inf	6 (0.00)	0.13	1.00	3.77	0.37 - 21.1
ATXN1	1 (0.02)	0 (0.00)	1.00	1.00	~ 0.4	0.01 - Inf	1 (0.00)	0.16	1.00	11.31	0.14 - 881.02
C9orf72	7 (0.12)	3 (0.12)	1.00	1.00	0.94	0.21 - 5.61	6 (0.00)	2.59×10^{-5}	8.56×10^{-4}	13.19	3.79 - 47.5
CCNF	1 (0.02)	0 (0.00)	1.00	1.00	~ 0.4	0.01 - Inf	10 (0.01)	0.61	1.00	1.13	0.03 - 7.95
CHMP2B	2 (0.03)	0 (0.00)	1.00	1.00	~ 0.8	0.08 - Inf	14 (0.01)	0.38	1.00	1.61	0.18 - 7.03
CYLD	1 (0.02)	0 (0.00)	1.00	1.00	~ 0.4	0.01 - Inf	3 (0.00)	0.29	1.00	3.77	0.07 - 46.94
DCTN1	2 (0.03)	1 (0.04)	1.00	1.00	0.80	0.04 - 47.33	9 (0.01)	0.22	1.00	2.51	0.26 - 12.15
EPHA4	2 (0.03)	0 (0.00)	1.00	1.00	~ 0.8	0.08 - Inf	2 (0.00)	3.54×10^{-2}	0.81	11.32	0.82 - 156.06
FIG4	12 (0.20)	4 (0.17)	1.00	1.00	1.20	0.36 - 5.12	53 (0.04)	5.69×10^{-3}	0.14	2.56	1.25 - 4.86

GLE1	2 (0.03)	0 (0.00)	1.00	1.00	~ 0.8	0.08 - Inf	26 (0.02)	1.00	1.00	0.87	0.1 - 3.48
MTHFR	2 (0.03)	1 (0.04)	1.00	1.00	0.80	0.04 - 47.33	21 (0.02)	0.71	1.00	1.08	0.12 - 4.41
PARK7	1 (0.02)	0 (0.00)	1.00	1.00	~ 0.4	0.01 - Inf	3 (0.00)	0.29	1.00	3.77	0.07 - 46.9
SS18L1	1 (0.02)	0 (0.00)	1.00	1.00	~ 0.4	0.01 - Inf	8 (0.01)	0.53	1.00	1.41	0.03 - 10.52
ZNF512B	2 (0.03)	0 (0.00)	1.00	1.00	~ 0.8	0.08 - Inf	4 (0.00)	$9.20x10^{-2}$	1.00	5.16	0.47 - 36.03

Table 2: Genes with pathogenic or likely pathogenic variants in cases or controls.
List of the 40 genes in which pathogenic (P) or likely pathogenic (LP) variants were found in cases and/or controls. On the left, the absolute number of ALS patients and controls of our cohort with minimal one P or LP variant in that gene is indicated along with the proportion of total cases (n=6013) and controls (n=2411) in percentage between brackets. P-value, adjusted p-value after correction for multiple testing, estimated odds ratio and 95% confidence interval are demonstrated in separate columns. To the right, the comparison with controls of the gnomAD database is made.

Abbreviations: Adj, adjusted; CI, confidence interval; Est, estimated; Inf, infinity; LP, likely pathogenic; NMHP, neuromuscular homepage; OR, odds ratio; P, pathogenic.

Discussion

In this paper, we present the genetic profile of a large sALS cohort. In total, 5.21% of patients carried a *C9orf72* repeat expansion and 2.68% a mutation in *SOD1*, *FUS, TARDBP, FUS, NEK1, OPTN* or *TBK1*. We observed potentially causative variants in other ALS-associated genes in an additional 3.26% of sALS patients (in 2.98% when patients carrying a combination with a *C9orf72* RE or with variants in *SOD1*, *FUS, TARDBP, NEK1, OPTN* and *TBK1* were excluded). In 17.33% we found isolated risk factors and in 71.83% no genetic variants at all were detected.

Since this cohort is enriched for patients with no clear molecular diagnosis, there could be an underrepresentation of pathogenic variants in the four most commonly mutated genes (*C9orf72, SOD1, FUS, TARDBP*). The strength of this project however is the characterization of genetic variation in the challenging group of unexplained sporadic patients, for whom the exact genetic architecture has not yet been unraveled. Indeed, until now, similar projects have mainly focused on fALS patients and we here describe the genetic landscape of a large cohort of sporadic ALS patients.[13] We believe that an overview of the genetic findings in sporadic ALS patients is the first step to a better understanding of the genetic architecture of ALS. Especially in the era of next generation sequencing and with the advent of gene-specific therapies, genetic screening is offered to an increasing number of patients. Therefore, a comprehensive genetic characterization of sALS patients is indispensable for informed counseling of these patients. Evidence on the strength of association of genes with ALS is also valuable for the development and analysis of gene panels.

Overall, when we analyzed genetic variation in 90 genes for which association with ALS has been reported, we observed an enrichment of P and LP sequencing variants in cases compared to controls. This enrichment did not remain significant when we excluded patients with known mutations, what might suggest that the difference between cases and controls is borne by established mutations. However, when we leave some less convincing genes out of the calculations, the difference is again clearly significant. This indicates that our list of genes is probably ‘polluted’ with genes that are not really associated with ALS and thus dilutes our results.

P and LP sequencing variants were detected in cases in only 38 of the 90 analyzed genes. In *NEK1*, *SOD1*, *FUS*, *OPTN* and *TARDBP* enrichment was significant. That these genes end up as top genes is not surprising since it is known that 2-3% of sALS patients carry strong pathogenic mutations in these genes.[5] In *SOD1*, we found mostly missense mutations spread throughout the gene. In *FUS* and *TARDBP*, variants reside mostly in the glycine-rich C-terminal or low complexity domains of the genes, as previously reported.[32] The association between (mainly loss-of-function (LOF)) variants in *NEK1* and ALS in this cohort has been described elsewhere and validates our approach.[33] In *OPTN*, bi-allelic and more rarely heterozygous variants have been described in ALS.[3]. In our cohort, four cases and no controls were homozygous for a variant in *OPTN*, the remaining variants were heterozygous (Fig. 1). That only *NEK1* and *SOD1* survived multiple testing is probably due to the low numbers of variants per individual gene. This strongly reduces statistical power, even though

this is the largest sALS cohort described to date. Using data from the gnomAD database to supplement our control data, all five originally significant genes were confirmed with higher significance levels. The lack of statistical power might also provide an explanation for the non-significant results in some of the 33 other genes with P and LP variants. This is illustrated by *TBK1*, an established ALS gene with non-significant enrichment of variants in our cohort, that was clearly significant when data from gnomAD were added. [34] For most of these genes without significant enrichment however, especially for the ones with odds ratios close to 1, we suspect that the probability of an association with ALS is rather low, as is illustrated as well by the more significant p-values when we leave some of these genes out of the calculations (see supra). The same is true for the 52 genes without P or LP variants detected in cases. Nevertheless, we cannot exclude that some of these genes harbor truly pathogenic mutations in only a small minority of ALS patients which are therefore not detected in our cohort. Genes conferring moderate risk for ALS are also not reliably identified by our pipeline, as is illustrated by the known polymorphism in *UNC13A.*

The yield of WGS (combined gene panel analysis and *C9orf72* repeat expansion detection in WGS data) in sALS patients was 11.09% overall and 7.87% when only taking *C9orf72* repeat expansions, *SOD1*, *FUS, TARDBP, NEK1, OPTN* and *TBK1* into consideration. This is lower than what is reported by others, but in those studies less stringent variant selection was performed (e.g. all novel or rare variants and/or ATXN2 intermediate expansions were reported in the yield) and a different number of genes was sequenced.[15,35] In 71.83% of sALS patients we did not find any P or LP variants so genetic determinants in this patient group remain unknown. Possible explanations are the presence of variants in not yet identified genes, variants in noncoding regions of the genome or epigenetic variation. Another possible reason is an oligogenic architecture where multiple less penetrant variants together constitute the risk of ALS in these patients. Such variants are not easily identified using the ACMG criteria. However, the fact that we found a significantly higher incidence of combinations in cases compared to controls could be a hint to a complex inheritance of (risk) variants in multiple genes.[8] Furthermore, the presence of P and LP variants in our control cohort points towards a reduced penetrance, which also fits with the oligogenic model. Even in *SOD1*, we found 2 controls carrying a pathogenic variant (0.08%), which is higher than the expected frequency (0.0025%) in controls assuming a lifetime risk for ALS of 1/400 and a frequency of 1% *SOD1* mutations in sporadic patients. [36] This illustrates that the penetrance at the population level is lower than what is expected based on smaller clinical cohorts, as has been shown for other neurodegenerative disorders before.[37,38] Probably other genetic and/or environmental factors play a role and explain why some mutation carriers seem more resistant to developing a clinical phenotype, again consistent with an oligogenic hypothesis. That we found a higher frequency of variants in controls (3.15%) compared to cases (2.98%) in the remaining genes after excluding

repeat expansions in *C9orf72* and variants in *SOD1*, *FUS, TARDBP, NEK1, OPTN* and *TBK1* – although the difference was not significant- can have multiple explanations: not all genes are robustly associated with ALS, some variants have a low penetrance or there is an overrepresentation of some variants because of populations stratification in a small control cohort. Further exploration of the penetrance of variants and the meaning of the presence of variants in the control population is of great importance since this dramatically influences genetic counseling in patients.

Of the unique P and LP variants we identified in cases, 60 (23.0%) were known ALS-associated mutations. With our data, we provide strong arguments that these known variants are truly pathogenic. The opposite should be interpreted with caution: since we don't take functional evidence or segregation data into account, not-confirmed variants could still be truly pathogenic and associated with ALS.

One of the problems in the analysis and interpretation of variants in complex genetic disorders as ALS is the lack of adequate scoring systems to assess the pathogenicity of variants. Frequently used criteria such as the ACMG guidelines are designed for Mendelian disorders and less applicable for diseases where variants might have low penetrance, be risk factors or might only be pathogenic in certain combinations.[17] Furthermore, these criteria are often vaguely defined and standard algorithms for implementing them are lacking, which leads to leeway and subjectivity in their practical implementation.[17,39] This makes an automated implementation on large datasets challenging. Since better classification systems are lacking, there is no consensus on how to assess the pathogenicity for variants in complex disorders and different methods are used through literature.[12,13] For this project, we used the software tool CharGer applying modified ACMG-criteria.[27] Despite our modifications, our criteria still favor the identification of more deleterious variants (e.g. nonsense and truncating variants) that are more often encountered in Mendelian disorders. On the other hand, some of the identified variants are equally present in cases and controls. This underlines the need for better classification systems for variants in complex diseases.

There are some limitations for this study. Since project MinE is a multicenter study, our cohort is composed of samples from different countries. Patients are mostly from European descent, which makes it a good reference work for this population, but caution is needed for the extrapolation to other populations. Although the control population was matched for the different populations included, the size of the control population was too small to calculate robust odd's ratios for every single variant. The randomly selected genes were not matched in properties to the ALS-associated genes. As brain-expressed genes tend to be longer, this could induce a bias because the longer the gene, the higher the chance of finding variants. [40] Also, since for the randomly selected genes no

mutations are known to be associated with ALS, some ACMG criteria are per definition not applicable to these genes.

In conclusion, this paper constitutes an inventory of pathogenic variation in a large cohort of sporadic ALS patients, which can serve as a valuable resource for genetic counseling and for the design of ALS gene panels. Not for all genes we could provide evidence for association with ALS. However, further research is needed to exclude very rare pathogenic variants or variants with moderate risk factors in these genes. To better explore the oligogenic hypothesis, even larger cohorts are needed to reach acceptable statistical power. A challenging journey lies ahead in order to obtain reliable comprehensive gene testing and counseling of sALS patients. However, it is a priority for clinical trial design and therapy development for ALS.

References

1. Hardiman O, Al-Chalabi A, Chio A, Corr EM, Logroscino G, Robberecht W, et al. Amyotrophic lateral sclerosis. Nat Rev Dis Primers 2017;3:17085.
2. Volk AE, Weishaupt JH, Andersen PM, Ludolph AC, Kubisch C. Current knowledge and recent insights into the genetic basis of amyotrophic lateral sclerosis. Medizinische Genetik 2018;30:252–8. https://doi.org/10.1007/s11825-018-0185-3.
3. Renton AE, Chiò A, Traynor BJ. State of play in amyotrophic lateral sclerosis genetics. Nat Neurosci 2014;17:17–23.
4. Byrne S, Heverin M, Elamin M, Bede P, Lynch C, Kenna K, et al. Aggregation of neurologic and neuropsychiatric disease in amyotrophic lateral sclerosis kindreds: a population-based case-control cohort study of familial and sporadic amyotrophic lateral sclerosis. Ann Neurol 2013;74:699–708.
5. Van Damme P, Robberecht W, Van Den Bosch L. Modelling amyotrophic lateral sclerosis: progress and possibilities. Disease Models & Mechanisms 2017;10:537–49. https://doi.org/10.1242/dmm.029058.
6. Al-Chalabi A, Fang F, Hanby MF, Leigh PN, Shaw CE, Ye W, et al. An estimate of amyotrophic lateral sclerosis heritability using twin data. J Neurol Neurosurg Psychiatry 2010;81:1324–6.
7. Leblond CS, Kaneb HM, Dion PA, Rouleau GA. Dissection of genetic factors associated with amyotrophic lateral sclerosis. Exp Neurol 2014;262 Pt B:91–101.
8. van Blitterswijk M, van Es MA, Hennekam EAM, Dooijes D, van Rheenen W, Medic J, et al. Evidence for an oligogenic basis of amyotrophic lateral sclerosis. Hum Mol Genet 2012;21:3776–84.
9. Al-Chalabi A, van den Berg LH, Veldink J. Gene discovery in amyotrophic lateral sclerosis: implications for clinical management. Nat Rev Neurol 2017;13:96–104.
10. He J, Mangelsdorf M, Fan D, Bartlett P, Brown MA. Amyotrophic Lateral Sclerosis Genetic Studies: From Genome-wide Association Mapping to Genome Sequencing. Neuroscientist 2015;21:599–615.
11. Alonso A, Logroscino G, Jick SS, Hernán MA. Incidence and lifetime risk of motor neuron disease in the United Kingdom: a population-based study. Eur J Neurol 2009;16:745–51.
12. Morgan S, Shatunov A, Sproviero W, Jones AR, Shoai M, Hughes D, et al. A comprehensive analysis of rare genetic variation in amyotrophic lateral sclerosis in the UK. Brain 2017;140:1611–8. https://doi.org/10.1093/brain/awx082.

13. Müller K, Brenner D, Weydt P, Meyer T, Grehl T, Petri S, et al. Comprehensive analysis of the mutation spectrum in 301 German ALS families. Journal of Neurology, Neurosurgery & Psychiatry 2018;89:817–27. https://doi.org/10.1136/jnnp-2017-317611.
14. Debray S, Race V, Crabbé V, Herdewyn S, Matthijs G, Goris A, et al. Frequency of C9orf72 repeat expansions in amyotrophic lateral sclerosis: a Belgian cohort study. Neurobiol Aging 2013;34:2890.e7–2890.e12.
15. Couthouis J, Raphael AR, Daneshjou R, Gitler AD. Targeted exon capture and sequencing in sporadic amyotrophic lateral sclerosis. PLoS Genet 2014;10:e1004704.
16. Project Mine ALS Sequencing Consortium. Project MinE: study design and pilot analyses of a large-scale whole-genome sequencing study in amyotrophic lateral sclerosis. European Journal of Human Genetics 2018;26:1537–46. https://doi.org/10.1038/s41431-018-0177-4.
17. Richards S, ; on behalf of the ACMG Laboratory Quality Assurance Committee, Aziz N, Bale S, Bick D, Das 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. Genetics in Medicine 2015;17:405–23. https://doi.org/10.1038/gim.2015.30.
18. Moisse M, Zwamborn RAJ, van Vugt J, van der Spek R, van Rheenen W, Kenna B, et al. The Effect of SMN Gene Dosage on ALS Risk and Disease Severity. Ann Neurol 2021;89:686–97.
19. Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD) [Internet] 2021. https://omim.org/.
20. Pestronk A. Neuromuscular homepage-ALS 2021. https://neuromuscular.wustl.edu/.
21. Benarroch L, Bonne G, Rivier F, Hamroun D. The 2021 version of the gene table of neuromuscular disorders (nuclear genome). Neuromuscul Disord 2020;30:1008–48.
22. Abel O, Powell JF, Andersen PM, Al-Chalabi A. ALSoD: A user-friendly online bioinformatics tool for amyotrophic lateral sclerosis genetics. Hum Mutat 2012;33:1345–51.
23. GTEx Portal 2021. https://gtexportal.org/home/.
24. Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, et al. Ensembl 2020. Nucleic Acids Res 2020;48:D682–8.
25. Dolzhenko E, Deshpande V, Schlesinger F, Krusche P, Petrovski R, Chen S, et al. ExpansionHunter: a sequence-graph-based tool to analyze variation in short tandem repeat regions. Bioinformatics 2019;35:4754–6.
26. van Es MA, Veldink JH, Saris CGJ, Blauw HM, van Vught PWJ, Birve A, et al. Genome-wide association study identifies 19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. Nat Genet 2009;41:1083–7.
27. Scott AD, Huang K-L, Weerasinghe A, Mashl RJ, Gao Q, Martins Rodrigues F, et al. CharGer: clinical Characterization of Germline variants. Bioinformatics 2019;35:865–7.
28. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res 2018;46:D1062–7.
29. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 2020;581:434–43.
30. Ou J, Zhu LJ. trackViewer: a Bioconductor package for interactive and integrative visualization of multi-omics data. Nat Methods 2019;16:453–4.
31. Brooks BR, Miller RG, Swash M, et al. El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. Amyotroph Lateral Scler Other Motor Neuron Disord 2000;1(5):293-9.
32. Amador M-D-M, Muratet F, Teyssou E, Boillée S, Millecamps S. New advances in Amyotrophic Lateral Sclerosis genetics: Towards gene therapy opportunities for familial and young cases. Rev Neurol 2021;177:524–35.
33. Kenna KP, van Doormaal PTC, Dekker AM, Ticozzi N, Kenna BJ, Diekstra FP, et al. NEK1 variants confer susceptibility to amyotrophic lateral sclerosis. Nat Genet 2016;48:1037–42.
34. Freischmidt A, Wieland T, Richter B, Ruf W, Schaeffer V, Müller K, et al. Haploinsufficiency of TBK1 causes familial ALS and fronto-temporal dementia. Nat Neurosci 2015;18:631–6.

35. Cady J, Allred P, Bali T, Pestronk A, Goate A, Miller TM, et al. Amyotrophic lateral sclerosis onset is influenced by the burden of rare variants in known amyotrophic lateral sclerosis genes. Ann Neurol 2015;77:100–13.
36. Pasinelli P, Brown RH. Molecular biology of amyotrophic lateral sclerosis: insights from genetics. Nat Rev Neurosci 2006;7:710–23.
37. Kay C, Collins JA, Miedzybrodzka Z, Madore SJ, Gordon ES, Gerry N, et al. Huntington disease reduced penetrance alleles occur at high frequency in the general population. Neurology 2016;87:282–8.
38. Gardiner SL, Boogaard MW, Trompet S, de Mutsert R, Rosendaal FR, Gussekloo J, et al. Prevalence of Carriers of Intermediate and Pathological Polyglutamine Disease-Associated Alleles Among Large Population-Based Cohorts. JAMA Neurol 2019;76:650–6.
39. Li Q, Wang K. InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. Am J Hum Genet 2017;100:267–80.
40. Zylka MJ, Simon JM, Philpot BD. Gene length matters in neurons. Neuron 2015;86(2):353-5.

Supplementary data

Supplementary tables

		Linked with ALS in				Mechanism of disease		
HUGO symbol	**Transcript used**	**Omim**	**NMHP**	**Gene-table**	**ALSoD**	**LOF**	**MIS**	**TRUNC**
ACSL5	ENST00000356116	N	Y	N	N	N	N	N
ALS2	ENST00000264276	Y	Y	Y	N	Y	Y	N
ANG	ENST00000336811	Y	Y	Y	N	Y	Y	N
ANXA11	ENST00000438331	Y	Y	Y	Y	Y	Y	N
APEX1	ENST00000216714	N	Y	N	N	N	N	N
APOE	ENST00000252486	N	Y	N	N	N	N	N
ATP13A2	ENST00000326735	N	Y	N	N	Y	N	N
ATXN1	ENST00000244769	N	Y	N	Y	Y	N	N
ATXN2	ENST00000550104	Y	Y	Y	N	N	N	N
BSCL2	ENST00000360796	N	Y	N	N	Y	N	N
C19orf12	ENST00000323670	N	Y	N	N	Y	N	N
C9orf72	ENST00000619707	Y	Y	Y	Y	N	Y	N
CACNA1H	ENST00000348261	N	Y	N	N	Y	N	N
CCNF	ENST00000397066	Y	Y	N	Y	N	Y	N
CFAP410	ENST00000397956	N	Y	N	Y	N	N	N
CHCHD10	ENST00000401675	Y	Y	Y	Y	Y	Y	N
CHMP2B	ENST00000263780	Y	Y	Y	N	Y	Y	N
CHRNA3	ENST00000326828	N	Y	N	N	Y	N	N
CHRNA4	ENST00000370263	N	Y	N	N	Y	N	N

CHRNB4	ENST00000261751	N	Y	N	N	Y	N	N
CNTF	ENST00000361987	N	Y	N	N	N	N	N
CX3CR1	ENST00000358309	N	Y	N	N	N	N	N
CYLD	ENST00000427738	Y	Y	N	N	Y	N	N
DAO	ENST00000228476	N	Y	N	N	N	N	N
DCTN1	ENST00000361874	Y	Y	Y	N	Y	Y	N
DCTN2	ENST00000434715	N	Y	N	N	N	N	N
DNAJC7	ENST00000457167	N	Y	N	N	N	N	N
DPP6	ENST00000377770	N	Y	N	N	N	N	N
ELP3	ENST00000256398	N	Y	N	N	N	N	N
EPHA4	ENST00000281821	N	Y	N	Y	Y	N	Y
ERBB4	ENST00000342788	Y	Y	Y	N	Y	Y	N
ERLIN1	ENST00000421367	N	Y	N	N	N	N	N
ERLIN2	ENST00000519638	N	Y	N	N	N	N	N
EWSR1	ENST00000414183	N	Y	N	N	Y	N	N
FGGY	ENST00000371218	N	Y	N	N	N	N	N
FIG4	ENST00000230124	Y	Y	Y	N	Y	Y	N
FUS	ENST00000254108	Y	Y	Y	Y	Y	Y	N
GLE1	ENST00000309971	N	Y	N	N	Y	N	Y
GLT8D1	ENST00000266014	N	Y	N	N	N	Y	N
GRIA2	ENST00000296526	N	Y	N	N	Y	N	N
GRN	ENST00000053867	N	Y	N	N	N	N	N
HFE	ENST00000357618	N	N	N	Y	N	N	N

HNRNPA1	ENST00000340913	Y	Y	Y	Y	Y	N	N
HNRNPA2B1	ENST00000354667	N	Y	N	N	Y	N	N
HSPB1	ENST00000248553	N	Y	N	N	N	N	N
HTT	ENST00000355072	N	Y	N	N	N	N	N
ITPR2	ENST00000381340	N	Y	N	N	Y	N	N
KIF5A	ENST00000455537	Y	Y	Y	Y	Y	N	N
LRP10	ENST00000359591	N	Y	N	N	Y	N	N
MATR3	ENST00000394805	Y	Y	Y	N	Y	Y	N
MTHFR	ENST00000376585	N	Y	N	N	Y	N	N
NEFH	ENST00000310624	Y	Y	Y	N	Y	N	Y
NEK1	ENST00000507142	Y	Y	Y	Y	Y	N	Y
NIPA1	ENST00000337435	N	Y	N	Y	N	N	N
OPTN	ENST00000378748	Y	Y	Y	Y	Y	Y	N
PARK7	ENST00000338639	N	Y	N	N	Y	N	N
PFN1	ENST00000225655	Y	Y	Y	Y	Y	Y	N
PNPLA6	ENST00000414982	N	Y	N	N	N	N	N
PON1	ENST00000222381	N	Y	N	N	N	N	N
PON2	ENST00000633531	N	Y	N	N	N	N	N
PON3	ENST00000265627	N	Y	N	N	N	N	N
PRPH	ENST00000257860	Y	Y	Y	N	Y	N	N
SCFD1	ENST00000458591	N	N	N	Y	N	N	N
SETX	ENST00000224140	Y	Y	Y	N	Y	Y	N

SIGMAR1	ENST00000277010	Y	Y	Y	N	Y	Y	N
SLC11A2	ENST00000394904	N	Y	N	N	N	N	N
SLC1A2	ENST00000278379	N	Y	N	N	Y	N	N
SMN1	ENST00000380707	N	Y	N	N	N	N	N
SMN2	ENST00000380743	N	Y	N	N	N	N	N
SOD1	ENST00000270142	Y	Y	Y	Y	Y	Y	N
SPG11	ENST00000261866	Y	Y	Y	N	Y	N	N
SQSTM1	ENST00000389805	Y	Y	Y	N	Y	Y	N
SS18L1	ENST00000331758	N	Y	N	N	Y	N	Y
SYNE1	ENST00000367255	N	Y	N	N	Y	N	Y
TAF15	ENST00000605844	N	Y	N	N	Y	N	N
TARDBP	ENST00000240185	Y	Y	Y	Y	Y	Y	N
TBK1	ENST00000331710	Y	Y	Y	Y	Y	Y	N
TIA1	ENST00000433529	Y	Y	N	N	N	N	N
TREM2	ENST00000373113	N	Y	N	N	N	N	N
TRPM7	ENST00000646667	Y	N	N	N	N	N	N
TTN	ENST00000589042	N	Y	N	N	N	N	N
TUBA4A	ENST00000248437	Y	Y	Y	Y	Y	Y	N
UBQLN2	ENST00000338222	Y	Y	Y	Y	Y	Y	N
UNC13A	ENST00000519716	N	Y	N	Y	N	N	N
VAPB	ENST00000475243	Y	Y	Y	Y	N	Y	N
VCP	ENST00000358901	Y	Y	Y	Y	Y	Y	N

VEGFA	ENST00000372055	N	Y	N	N	N	N	N
VRK1	ENST00000216639	N	Y	N	N	N	N	N
ZDHHC6	ENST00000369405	N	Y	N	N	N	N	N
ZNF512B	ENST00000369888	N	Y	N	N	Y	N	N

Supplementary table 1: List of ALS-associated genes
List of ALS-associated genes (HUGO symbols[1]) and the transcripts used. For each gene is indicated in which of the four databases (Omim[2], NMHP [3], gene table[4], ALSoD[5]) the gene is linked with ALS. Additionally, for each gene it is specified whether loss-of-function (LOF) was considered as a known mechanism of disease (genes for which PVS1 could be triggered), whether missense variants (MIS) were a mechanism of disease (genes for which PP2 could be triggered) and whether primarily truncating variants (TRUNC) caused disease (genes for which BP1 could be triggered). More information on how the mechanism of disease was determined, can be found in supplementary figure 1.

Abbreviations: LOF, loss-of-function; MIS, missense; N, No; NMHP, neuromuscular homepage; TRUNC, truncating; Y, Yes.

Gene	**Transcript used**	**LOF**	**MIS**	**TRUNC**
ACSM2B	ENST00000329697	N	N	N
AMN	ENST00000299155	N	N	N
AQP11	ENST00000313578	N	N	N
ARG1	ENST00000356962	N	N	N
BCL2L15	ENST00000393316	N	N	N
BLNK	ENST00000224337	Y	N	N
C20orf85	ENST00000371168	N	N	N
CAPNS2	ENST00000457326	N	N	N
CCBE1	ENST00000649564	N	N	N
CCDC134	ENST00000255784	N	N	N
CCR6	ENST00000341935	N	N	N
CCR7	ENST00000246657	Y	N	N
CD109	ENST00000287097	N	N	N
CDH3	ENST00000264012	N	N	N
CLEC12A	ENST00000355690	N	N	N
CPB1	ENST00000491148	N	N	N

CTSW	ENST00000307886	N	N	N
CXCL6	ENST00000226317	N	N	N
CYP3A5	ENST00000222982	N	N	N
DBF4	ENST00000265728	N	N	N
DMGDH	ENST00000255189	N	N	N
DUS2	ENST00000565263	N	N	N
ENTPD7	ENST00000370489	N	N	N
EPHX3	ENST00000221730	N	N	N
F5	ENST00000367797	Y	N	N
FAM161A	ENST00000404929	N	N	N
FAM180A	ENST00000338588	Y	N	N
FAM83B	ENST00000306858	N	N	N
FANCI	ENST00000310775	N	N	N
FOXA3	ENST00000302177	Y	N	N
GBGT1	ENST00000372040	N	N	N
GBP6	ENST00000370456	N	N	N
GJC1	ENST00000592524	Y	N	N
GNL3L	ENST00000360845	N	N	N
GP2	ENST00000381362	Y	N	N
GRAP	ENST00000284154	N	N	N
GRHL3	ENST00000350501	N	N	N
HSH2D	ENST00000616645	N	N	N
IGHD3-10	ENST00000390583	N	N	N
IGHV6-1	ENST00000390593	N	N	N
IGKV1D-16	ENST00000492446	N	N	N
IKBIP	ENST00000299157	N	N	N
ISX	ENST00000308700	Y	N	N
KIFC1	ENST00000428849	N	N	N
KLHL8	ENST00000273963	N	N	N

KLRD1	ENST00000336164	N	N	N
KNL1	ENST00000346991	N	N	N
KNTC1	ENST00000333479	N	N	N
KRT72	ENST00000293745	Y	N	N
LCE1A	ENST00000335123	N	N	N
LCE2C	ENST00000368783	N	N	N
LRP12	ENST00000276654	Y	N	N
LYG1	ENST00000409448	N	N	N
MAB21L2	ENST00000317605	Y	N	N
MCOLN2	ENST00000370608	N	N	N
MIS18A	ENST00000290130	N	N	N
MLLT3	ENST00000380338	Y	N	N
MOXD1	ENST00000367963	N	N	N
MYB	ENST00000341911	N	N	N
NAT8	ENST00000272425	N	N	N
NEU4	ENST00000325935	N	N	N
NHSL1	ENST00000427025	N	N	N
NOD2	ENST00000300589	N	N	N
P2RX1	ENST00000225538	N	N	N
P2RX2	ENST00000343948	N	N	N
PADI3	ENST00000375460	N	N	N
PBK	ENST00000522944	N	N	N
PDE3A	ENST00000359062	N	N	N
PIPOX	ENST00000323372	N	N	N
PLA2G7	ENST00000274793	N	N	N
PRR15	ENST00000319694	N	N	N
PRSS12	ENST00000296498	N	N	N
PTGS1	ENST00000362012	N	N	N
PXMP4	ENST00000409299	N	N	N

RAB43	ENST00000315150	N	N	N
RASGRP4	ENST00000615439	N	N	N
RTP4	ENST00000259030	N	N	N
SERPINB2	ENST00000457692	Y	N	N
SH2D4A	ENST00000265807	N	N	N
SH3RF3	ENST00000309415	N	N	N
SIGLEC9	ENST00000440804	N	N	N
SPINK4	ENST00000379721	N	N	N
SPRR1B	ENST00000307098	N	N	N
SPRR2G	ENST00000368748	N	N	N
ST6GALNAC1	ENST00000156626	N	N	N
SUV39H2	ENST00000354919	Y	N	N
TK1	ENST00000588734	N	N	N
TMEM86A	ENST00000280734	N	N	N
TNFSF13B	ENST00000375887	N	N	N

Supplementary table 2: List of random genes

List of random genes (HUGO symbols[1]) selected from the GTEx Portal[6] and the transcripts used. For each gene it is specified whether loss of function (LOF) was considered as a known mechanism of disease (genes for which PVS1 could be triggered), whether missense variants (MIS) were a mechanism of disease (genes for which PP2 could be triggered) and whether primarily truncating variants (TRUNC) caused disease (genes for which BP1 could be triggered). More information on how the mechanism of disease was determined, can be found in Supplementary figure 1.

Abbreviations: LOF, loss-of-function; MIS, missense; N, No; TRUNC, truncating; Y, Yes

Criteria	Description	Comment
PVS1	Null variant in a gene where LOF is a known mechanism of disease	Strategy used to determine a list of genes where LOF is a known mechanism of disease: see supplementary figure 1. See supplementary table 1 for this list of genes.
PS1	Same amino acid change as previously established pathogenic variant regardless of nucleotide change	Strategy used to determine a list of previously established pathogenic variants: see supplementary figure 1.
PS2	De novo in a patient with no family history	Not used since automated attribution is not possible.
PS3	Well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene (product)	Not used since automated attribution is not possible.
PS4	Prevalence of variant in affected individuals is significantly increased compared with the prevalence in controls	Not used.
PM1	Mutational hot spot and/or functional domain	Strategy used to delineate hot spot regions: see supplementary figure 1.
PM2	Absent from controls in population database	Does not need to be absent, maximum minor allele frequency threshold = 0.05%
PM3	For recessive disorders, detected in trans with a pathogenic variant	Not used since cis/trans status could not be reliably detected from our data.
PM4	Protein length changes as a result of in-frame deletions/insertions in a non-repeat region / stop-loss variants	Variants in repeat regions are excluded.[1]
PM5	Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before	Strategy used to determine a list of previously established pathogenic variants: see supplementary figure 1.
PM6	Assumed de novo, but without confirmation of paternity and maternity	Not used since automated attribution is not possible.
PP1	Cosegregation with disease in multiple affected family members in a gene definitively known to cause the disease	Not used since automated attribution is not possible.
PP2	Missense variant in a gene that has a low rate of benign missense variation and in which missense variants are a common mechanism of disease	Strategy used to determine a list of genes where missense variants are a known mechanism of disease: see supplementary figure 1. See supplementary table 1 for this list of genes.
PP3	Computational evidence supports a deleterious effect	Prediction tools used: fathmm-MKL_coding, MutationTaster, MutationAssessor, CADD, DEOGEN2, PrimateAI, BayesDel_addAF, LIST-S2, MVP, Eigen, M-CAP and SIFT and rf and ada for splice variants.

		Triggered when ⅔ (66%) of the tools with a clear call (deleterious or benign) predicted a deleterious effect.
PP4	Patient's phenotype/family history is highly specific for a disease with a single genetic etiology	Not used (not applicable in the oligogenic hypothesis).
PP5	Reputable source reports variant as pathogenic	Strategy used to determine a list of previously established pathogenic variants: see supplementary figure 1.
BA1	Allele frequency >5% in population databases	
BS1	Allele frequency greater than expected for disorder	Threshold used: 1/400.
BS2	Observed in a healthy adult individual for a recessive, dominant or X-linked disorder with full penetrance expected at an early age	Not used (not applicable for an adult-onset disorder like ALS).
BS3	Well-established in vitro or in vivo functional studies show no damaging effect on the gene (product)	Not used since automated attribution is not possible.
BS4	Lack of segregation in affected members of a family	Not used since automated attribution is not possible.
BP1	Missense variant in a gene for which primarily truncating variants are known to cause disease	Strategy used to determine a list of genes for which truncating variants are a known mechanism of disease: see supplementary figure 1. See supplementary table 1 for this list of genes.
BP2	Observed in trans with a pathogenic variant for a fully penetrant dominant gene/disorder or observed in cis with an pathogenic variant in any inheritance pattern	Not used since cis/trans status could not be reliably detected from our data.
BP3	In-frame deletions/insertions in a repetitive region without a known function	Repeat regions were defined in *ATXN1*, *ATXN2*, *NIPA1* and *HTT*.[1]
BP4	Computational evidence suggests no impact	Same prediction tools used as for PP3. Triggered when ⅔ (66%) of the tools with a clear call (deleterious or benign) predicted a benign effect.
BP5	Variant found in a case with an alternate molecular basis for disease	Not used (hypothesis of oligogenic architecture in ALS)
BP6	Reputable source reports variant as benign	Strategy used to determine a list of previously established benign variants: see supplementary figure 1.
BP7	Synonymous variant with no predicted impact on splicing and nucleotide is not highly conserved	Synonymous variants not present in dbscSNV database.[8]

Supplementary table 3: Overview of ACMG-criteria and their adaptations. Modified from Richards S. et al [7]

[1]Repeat regions were defined in *ATXN1*, *ATXN2*, *NIPA1* and *HTT*: *ATXN1*: chr6:16327633-16327725; *ATXN2* chr12:111598827-111599020; *NIPA1* chr15:22786665-22786705; *HTT*: chr4:3074876-3074936.

Abbreviations: ALS, amyotrophic lateral sclerosis; LOF, loss-of-function.

HUGO symbol	Benign	Likely benign	Uncertain significance	Likely pathogenic	Pathogenic
ALS2	26	18	71	6	11
ANG	11	7	8	0	8
ANXA11	0	0	5	1	4
ATP13A2	0	0	1	0	0
ATXN2	0	0	2	0	0
C9orf72	22	5	57	0	1
CACNA1H	0	0	2	0	0
CCNF	0	0	6	0	4
CFAP410	0	0	1	0	0
CHCHD10	8	15	40	1	3
CHMP2B	24	6	28	0	6
CHRNA3	0	0	1	4	0
CHRNA4	0	2	1	0	0
CHRNB4	2	0	1	0	0
CX3CR1	2	0	0	0	0
CYLD	0	0	1	0	1
DAO	0	1	3	1	0
DCTN1	36	52	196	1	1
DNAJC7	0	0	1	0	0
ELP3	0	0	1	0	0
EPHA4	0	0	0	0	1
ERBB4	1	1	5	3	2
ERLIN1	0	0	0	0	1
ERLIN2	0	0	4	0	0
EWSR1	0	0	6	0	0
FIG4	21	11	26	0	7
FUS	44	29	125	8	20
GLE1	0	0	2	0	2
GLT8D1	0	0	5	0	0
HNRNPA1	0	0	3	0	1
KIF5A	1	1	4	1	0

LRP10	0	0	1	0	0
MATR3	37	34	74	0	4
NEFH	9	1	3	0	0
NEK1	0	2	6	1	2
NIPA1	0	0	1	0	0
OPTN	25	15	56	2	14
PARK7	0	0	1	0	1
PFN1	0	0	1	0	4
PNPLA6	0	0	2	0	0
PON1	0	0	1	0	0
PON3	0	1	0	0	0
PRPH	0	1	2	0	0
SCFD1	0	0	2	0	0
SETX	96	85	285	3	11
SIGMAR1	5	11	39	7	4
SOD1	8	5	31	63	44
SPG11	5	2	14	2	14
SQSTM1	20	36	99	2	10
SS18L1	0	2	5	0	1
SYNE1	0	0	0	0	4
TAF15	0	0	7	0	1
TARDBP	14	10	66	3	16
TBK1	13	14	30	3	17
TIA1	0	0	7	1	1
TREM2	0	0	1	0	0
TRPM7	0	0	2	0	0
TUBA4A	0	0	4	0	6
UBQLN2	4	3	26	1	4
UNC13A	0	0	1	0	0
VAPB	51	30	143	0	2
VCP	26	30	82	3	16
VRK1	0	0	0	0	1

Supplementary table 4: Reclassification of ALS-associated variants from ClinVar and NMHP.
Variants associated with ALS in the 90 genes of interest were extracted from ClinVar and the NMHP.[3,9] Since the variants in ClinVar and the NMHP are not curated, we performed a reclassification of variants in an attempt to correct for possibly misattributed variants. The reclassification procedure is described in *supplementary figure 1*. In this table, the number of variants with a final reclassification as pathogenic, likely pathogenic, unknown significance, likely benign and benign are displayed for each gene.

Gene	Case (%)	Control (%)	P-value	Adj p-value	Est. OR	95% CI
FAM180A	0 (0)	1 (0.04)	$2.86x10^{-1}$	1.00	0	0 - 15.64
FOXA3	0 (0)	1 (0.04)	$2.86x10^{-1}$	1.00	0	0 - 15.64
MAB21L2	0 (0)	1 (0.04)	$2.86x10^{-1}$	1.00	0	0 - 15.64
MLLT3	1 (0.02)	1 (0.04)	$4.91x10^{-1}$	1.00	0.4	0.01 - 31.48
BLNK	3 (0.05)	0 (0)	$5.62x10^{-1}$	1.00	~ 1.2	0.17 - Inf
SERPINB2	4 (0.07)	0 (0)	$5.84x10^{-1}$	1.00	~1.6	0.26 - Inf
F5	6 (0.1)	1 (0.04)	$6.81x10^{-1}$	1.00	2.41	0.29 - 110.72
ISX	9 (0.15)	2 (0.08)	$7.39x10^{-1}$	1.00	1.81	0.37 - 17.19
GP2	14 (0.23)	4 (0.17)	$7.94x10^{-1}$	1.00	1.4	0.44 - 5.86
KRT72	2 (0.03)	0 (0)	1.00	1.00	~ 0.8	0.08 - Inf
LRP12	1 (0.02)	0 (0)	1.00	1.00	~ 0.4	0.01 - Inf

Supplementary table 5: Randomly selected genes with pathogenic or likely pathogenic variants in cases or controls.
List of the 11 genes in which pathogenic (P) or likely pathogenic (LP) variants were found in cases and/or controls. The absolute number of ALS patients and controls of our cohort with minimal one P or LP variant in that gene is indicated along with the proportion of total cases (n=6013) and controls (n=2411) in percentage between brackets. P-value, adjusted p-value after correction for multiple testing, estimated odds ratio and 95% confidence interval are demonstrated in separate columns.

Abbreviations: Adj, adjusted; CI, confidence interval; Est, estimated; Inf, infinity; LP, likely pathogenic; NMHP, neuromuscular homepage; OR, odds ratio; P, pathogenic.

Variant 1	Variant 2	Variant 3	Number of cases	Number of controls
C9orf72 RE	*CACNA1H*	-	3	0
C9orf72 RE	*SPG11*	-	2	0
C9orf72 RE	*ITPR2*	-	1	0
C9orf72 RE	*OPTN*	-	1	0
C9orf72 RE	*SYNE1*	-	1	0
C9orf72 RE	*BSCL2*	-	1	0
C9orf72 RE	*CHMP2B*	-	1	0
C9orf72 RE	*FIG4*	-	1	0
C9orf72 RE	*ATXN2* RE	-	4	0
C9orf72 RE	*NIPA1* RE	-	5	1
C9orf72 RE	*UNC13A*	-	40	2
SOD1	*ZNF512B*	-	1	0
SOD1	*ATXN2* RE	-	1	0
SOD1	*NIPA1* RE	-	1	0
SOD1	*UNC13A*	-	1	0
FUS	*SETX*	-	1	0
FUS	*UNC13A*	-	1	0
TARDBP	*UNC13A*	-	1	0
ALS2	*UNC13A*	-	1	0
ANG	*UNC13A*	-	1	0
ANXA11	*UNC13A*	-	1	1
ATP13A2	*UNC13A*	-	1	0
C9ORF72 (SV)	*UNC13A*	-	0	1
CACNA1H	*UNC13A*	-	3	3
CCNF	*UNC13A*	-	1	0
CHRNA3	*UNC13A*	-	1	0
CHRNA4	*UNC13A*	-	1	0

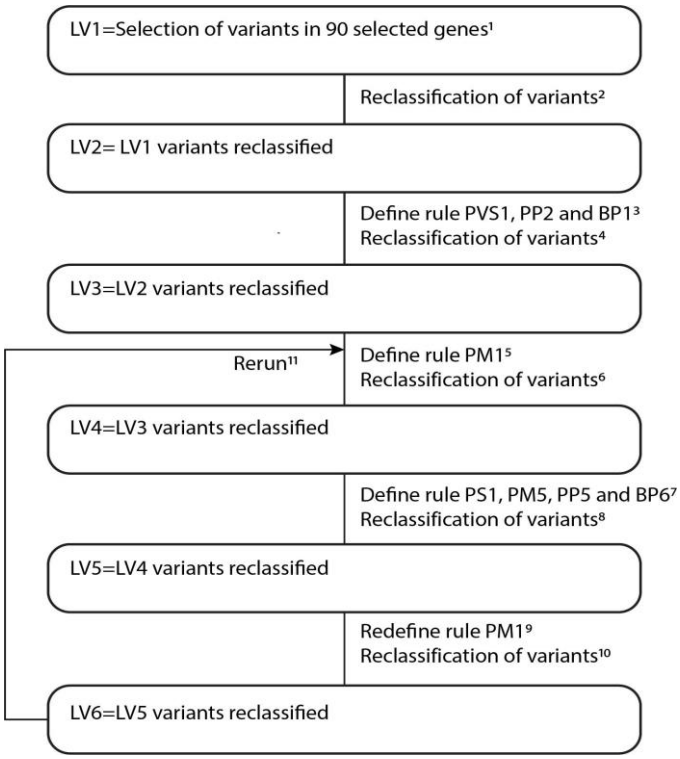
Variant 1	Variant 2	Variant 3	Number of cases	Number of controls
FIG4	*UNC13A*	-	0	1
GLT8D1	*UNC13A*	-	1	0
NEFH	*UNC13A*	-	1	1
NEK1	*UNC13A*	-	7	0
OPTN	*UNC13A*	-	5	2
PRPH	*UNC13A*	-	2	0
SETX	*UNC13A*	-	2	1
TBK1	*UNC13A*	-	1	0
VCP	*UNC13A*	-	1	0
ZNF512B	*UNC13A*	-	1	0
C9ORF72 (SV)	*NIPA1* RE	-	1	0
C9ORF72 (SV)	*ATXN1* RE	-	1	0
CACNA1H	*ATXN2* RE	-	2	2
CHMP2B	*NIPA1* RE	-	1	0
NEFH	*ATXN2* RE	-	1	0
NEK1	*ATXN2* RE	-	2	0
NEK1	*NIPA1* RE	-	1	0
PRPH	*ATXN2* RE	-	1	0
SETX	*ATXN2* RE	-	1	0
SYNE1	*ATXN2* RE	-	2	0
SYNE1	*NIPA1* RE	-	1	0
TBK1	*ATXN2* RE	-	0	1
ANXA11	*CACNA1H*	-	1	0
BSCL2	*ITPR2*	-	1	0
CACNA1H	*CHRNA4*	-	1	0
PRPH	*NEK1*	-	1	0
SPG11	*NEFH*	-	1	0

Variant 1	Variant 2	Variant 3	Number of cases	Number of controls
SS18L1	*FIG4*	-	1	0
CHRNA3	*CHRNA3*	-	0	1
ALS2	*NEFH*	-	0	1
ATXN2 RE	*UNC13A*	-	22	7
NIPA1 RE	*UNC13A*	-	22	9
ATXN1 RE	*NIPA1* RE	-	1	0
ATXN2 RE	*NIPA1* RE	-	1	0
C9orf72 RE	*CHRNA3*	*SPG11*	1	0
C9orf72 RE	*SYNE1*	*NIPA1* RE	1	0
C9orf72 RE	*CACNA1H*	*UNC13A*	1	0
C9orf72 RE	*SETX*	*UNC13A*	1	0
C9orf72 RE	*ATXN2* RE	*UNC13A*	2	0
FUS	*NIPA1* RE	*UNC13A*	1	0
C9ORF72 (SV)	*C9ORF72* (SV)	*C9ORF72* (SV)	0	1
GLT8D1	*SYNE1*	*UNC13A*	1	0
SPG11	*CACNA1H*	*UNC13A*	1	0
SPG11	*NIPA1* RE	*UNC13A*	0	1

Supplementary table 6: Combinations of variants.

Abbreviations: RE, repeat expansion; SV, sequencing variant.

Supplementary figures



Supplementary figure 1: Workflow to define rule PVS1, PS1, PM1, PM5, PP2, PP5, BP1 and BP6 and to reclassify ALS-associated variants

To define rule PVS1, PS1, PM1, PM5, PP2, PP5, BP1 and BP6 we started from a list of established variants associated with ALS in the 90 genes of interest. These variants were selected from 2 databases, ClinVar and the NMHP, as described below (see: create LV1).[3,9] Subsequently, we used the workflow as illustrated in this figure to define the individual ACMG rules (PVS1, PS1, PM1, PM5, PP2, PP5, BP1 and BP6) and to repeatedly obtain reclassifications of these established variants. The consecutive steps were repeated until all variants obtained a 'stable' reclassification. From this final classification, the pathogenic (P) and likely pathogenic (LP) were considered 'known ALS-associated mutations' (see main paper).

[1]<u>Create LV1: list of established variants in the 90 selected genes</u>

To create a list (LV1) of established P, LP, benign (B) and likely benign (LB) variants and variants of unknown significance (VUS) in the 90 genes of interest, we selected following variants:

- in ClinVar, we selected variants associated with ALS and classified as P, LP, VUS, LB and B in the 90 genes of interest. When conflicting interpretations were submitted in ClinVar, we classified the variant as pathogenic when we found P, LP and/or VUS submissions; as benign if we found B, LB and/or VUS submissions; and as VUS if both P/LP and B/LB submissions were present for the variant.
- in the NMHP, we selected variants associated with ALS (no exact classification is provided in this database)

[2]<u>Reclassification of variants (1)</u>

We ran CharGer with our modified ACMG-criteria on LV1, omitting rule PVS1, PM1, PP2, PS1, PM5, PP5, BP1 and BP6; thereby obtaining CharGer scores for all variants (the more positive the score the more evidence for pathogenicity) as well as an independent classification as P, LP, VUS, LB or B according to our preliminary ACMG rules.
Subsequently, we reclassified the variants of LV1 as follows, creating LV2:

- 'Pathogenic' or 'Likely pathogenic'
    - Variants reported as P or LP in ClinVar *AND* classified as P, LP or VUS according to our preliminary ACMG-criteria *AND* a positive CharGer score.
    - Variants from the NMHP that were classified as P or LP according to our preliminary ACMG classification
- Variants of unknown significance with pathogenic evidence (VUS)+
    - VUS variants (classified as VUS in ClinVar or after using the preliminary ACMG-criteria for the variants of the NMHP) for which the CharGer total score was >2. The rationale for this cutoff is illustrated in supplementary figure 2.
- Variants of unknown significance (VUS)
    - Variants reported as P or LP in ClinVar *AND* a negative CharGer score

    - Variants reported as VUS in ClinVar *AND* classified as VUS according to our preliminary ACMG-criteria
    - Variants from the NMHP reported as VUS according to our preliminary ACMG-criteria
- 'Likely benign' or 'Benign'
    - Variants reported as B or LB in ClinVar *AND* classified as B, LB or VUS according to our preliminary ACMG-criteria
    - Variants from the NMHP that were classified as B or LB according to our preliminary ACMG criteria

[3]Defining rule PVS1, PP2 and BP1
1) Define rule PVS1: create a list of genes for which loss of function (LOF) variants are a known mechanism of disease
From LV2, we selected variants reclassified as P, LP or VUS+ as 'ALS-associated variants'.
We summarized per gene how many of these ALS-associated variants were identified, how many of them were missense variants and how many were LOF variants.
We defined 'genes for which LOF is a known mechanism of disease' as genes in which minimal one ALS-associated LOF variant was present. Additionally, genes with a LoFtool score of 0.25 or more were added to complete this list.[10]

2) Define rule PP2: create a list of genes for which missense variants are a known mechanism of disease
'Genes for which missense variants were considered a mechanism of disease' were defined as genes in which minimal 5 ALS-associated missense variants were found and where the percentage of ALS-associated missense variants was more than 25% of all pathogenic variants in that gene.

3) *Define rule BP1: create a list of genes with primarily truncating variants as mechanism of disease*
We defined these genes as genes in which minimal one ALS-associated LOF variant was present and in which no ALS-associated missense variants were present.

[4]Reclassification of variants (2)
Same procedure as stated above (Reclassification of variants (1)) but including rule PVS1 and PP2 when running CharGer. Rule PM1, PS1, PM5, PP5 and BP6 were still omitted.
Obtaining updated CharGer scores and preliminary ACMG classifications, we used the same reclassification rules as stated above to select variants with a reclassification as P, LP and VUS+, creating LV3.

[5]Defining rule PM1: delineate hotspots regions
Starting from LV3, a sliding window of 51 nucleotides was used on the genomic regions to determine regions in which minimal 4 variants classified as P, LP or VUS+ (in LV3) were present. These regions were labeled as mutational hot spots.

[6]Reclassification of variants (3)
Same procedure as stated above (Reclassification of variants (1)) but including rule PVS1, PP2, PM1 and BP1 when running CharGer. Rule PS1, PM5, PP5 and BP6 were still omitted.
Obtaining updated CharGer scores and preliminary ACMG classifications, we used the same reclassification rules as stated above to select variants with a reclassification as P, LP and VUS+, creating LV4.

[7]Defining rule PS1, PM5, PP5 and BP6 (same amino acid change as previously established pathogenic variant/novel missense change at amino acid residue where a different missense change reported as pathogenic has been seen before/reputable source reports variant as pathogenic/benign)
Reported pathogenic variants were defined as the variants from LV4 with a reclassification as P or LP. Reported benign variants were defined as the variants from LV4 with a reclassification as B or LB.

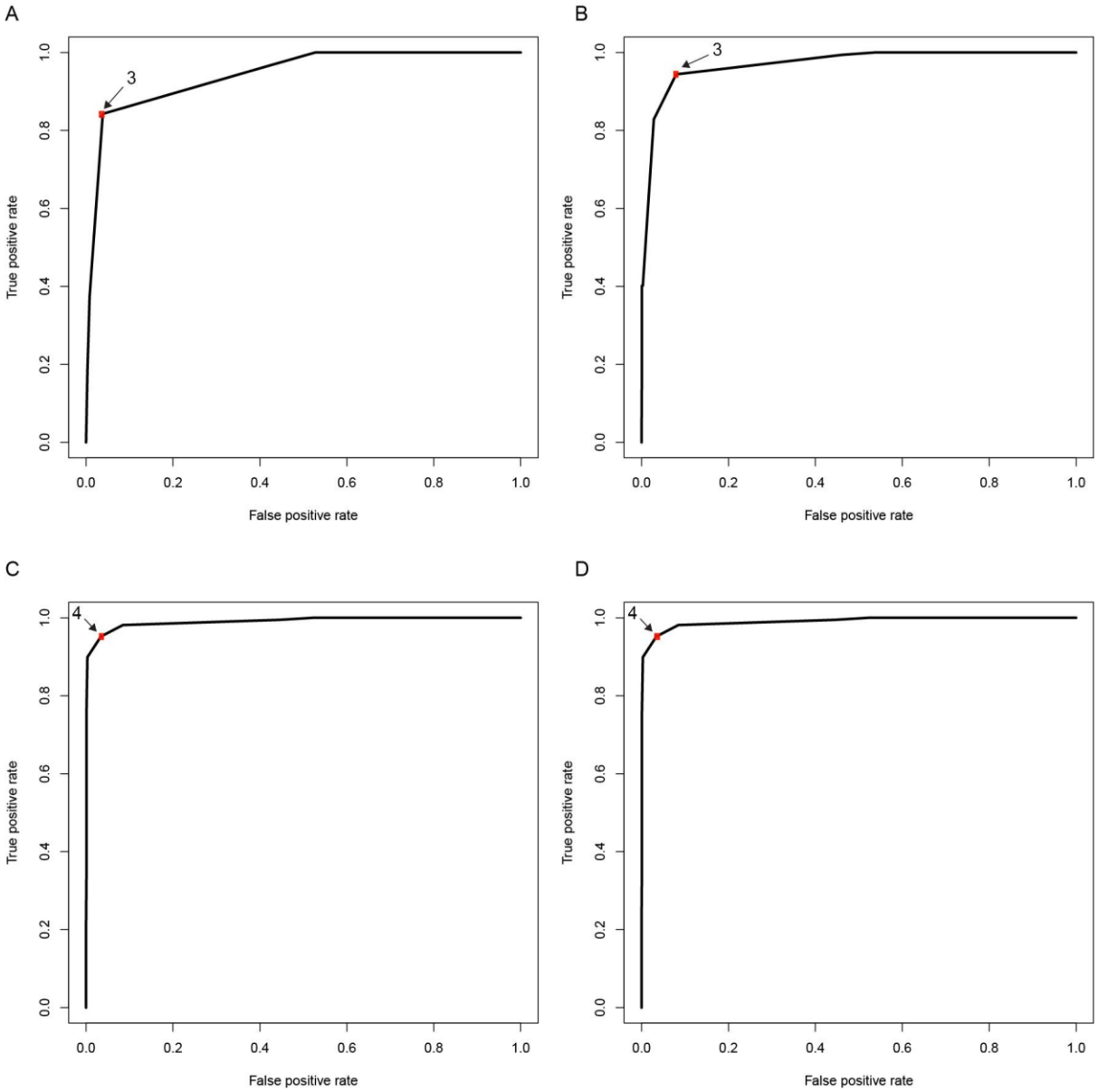
[8]Reclassification of variants (4)
Same procedure as stated above (Reclassification of variants (1)) but including rule PVS1, PP2, PM1, PS1, PM5, PP5, BP1 and BP6 when running CharGer. Obtaining updated CharGer scores and preliminary ACMG classifications, we used the same reclassification rules as stated above to select variants with a reclassification as P, LP and VUS+, creating LV5.

[9]Redefine rule PM1
Defining new hot spot regions using the same procedure as above (defining rule PM1: delineate hotspots regions), but starting from LV5.

[10]Reclassification of variants (5)
Same procedure as stated above (Reclassification of variants (1)) but including rule PVS1, PP2, PM1, PS1, PM5, PP5 when running CharGer. Obtaining updated CharGer scores and preliminary ACMG classifications, we used the same reclassification rules as stated above to select variants with a reclassification as P, LP and VUS+, creating LV6.

[11]Rerun
The consecutive steps were repeated until all variants obtained a 'stable' classification. The VUS+ cutoff was recalculated based on a roc curve analysis as explained in supplementary figure 2.

Abbreviations: B, benign; LB, likely benign; LOF, loss of function; LP, likely pathogenic; LV, list of variants; NMHP, Neuromuscular homepage; P, pathogenic; VUS, variant of unknown significance; VUS+, variant of unknown significance with pathogenic evidence



Supplementary figure 2: ROC curve analysis of the CharGer-scores for predicting ACMG-class (benign-likely, benign vs pathogenic-likely, pathogenic) in ALS-associated variants.

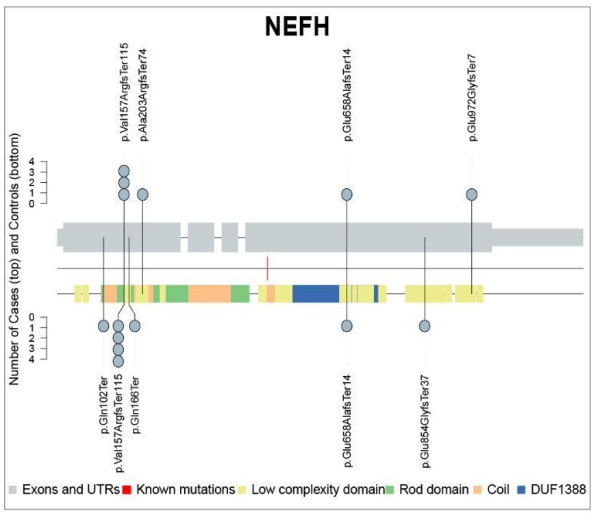
ROC curve analysis of the CharGer-scores for predicting ACMG-class (as reported in ClinVar or as calculated for the NMHP variants according to our modified ACMG-criteria).

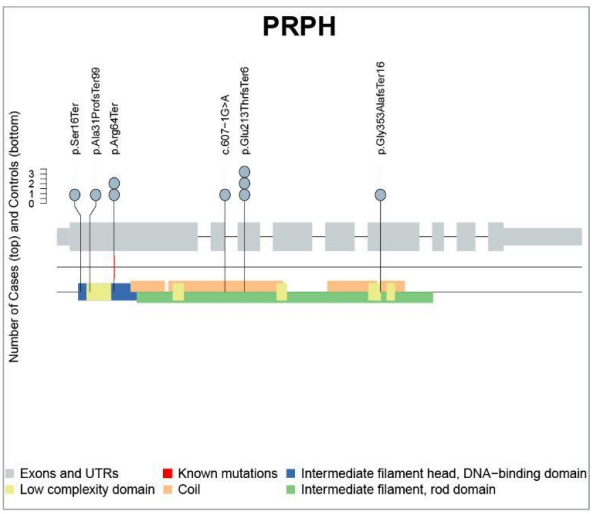
A: ACMG-class was calculated by running CharGer but omitting rules: PVS1, PM1, PP2, PS1, PM5, PP5, BP1 and BP6. The optimal cut-off value was found at a CharGer score of 3 (>2). AUC:0.9433; 95% CI [0.9298-0.9568]. Thus, VUS variants with a CharGer score of >2 were considered as VUS+ variants and included in the analysis for the selection of LOF genes and missense genes *(see supplementary figure 1)*.

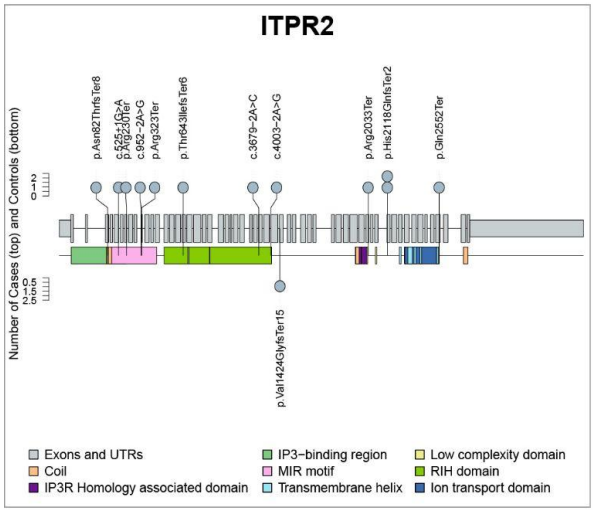
B: ACMG-class was calculated, omitting rule PM1 and PP5. The optimal cut-off value was found at a CharGer score of 3 (>2). AUC: 0.9702; 95% CI [0.9614-0.979]). VUS-variants with a CharGer-score of >2 were considered VUS+ variants and included for analysis for the delineation of hot spots *(see supplementary figure 1)*.

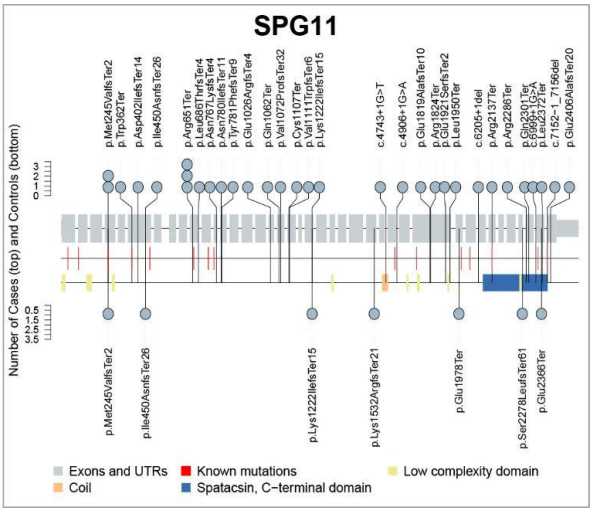
C: ACMG-class was calculated with all the rules. The optimal cut-off value was found at a CharGer score of 4 (>3). AUC: 0.9904; 95% CI [0.9853-0.9955]). VUS-variants with a CharGer-score of >3 were considered VUS+ variants *(see supplementary figure 1)*.

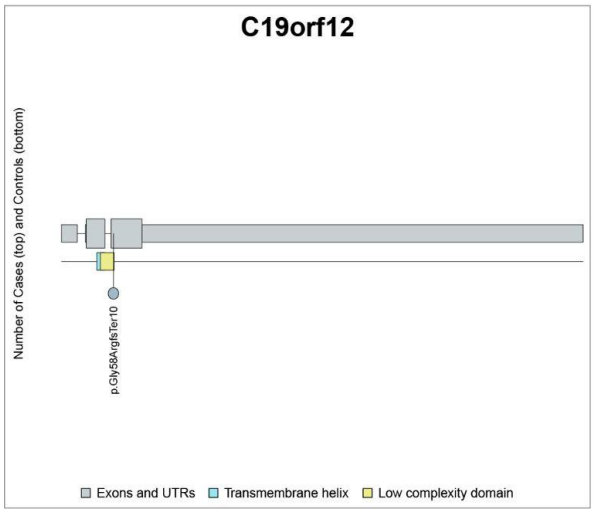
D: The final ACMG-class was calculated with all the rules. The optimal cut-off value was found at a CharGer score of 4 (>3). AUC: 0.9904; 95% CI [0.9853-0.9955]). VUS-variants with a CharGer-score of >3 were considered VUS+ variants *(see supplementary figure 1)*.

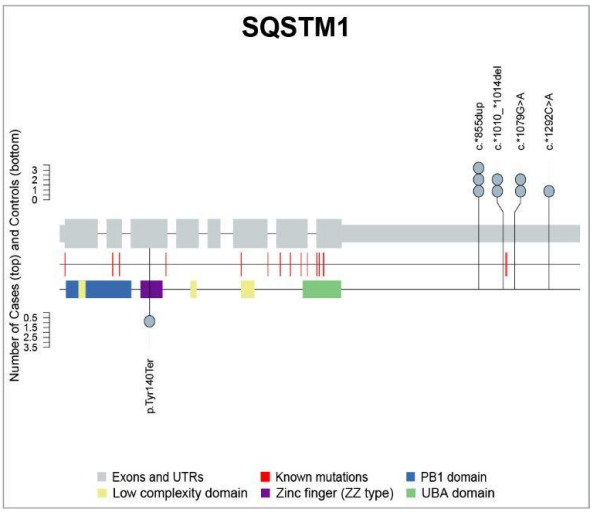

NEFH
Number of Cases (top) and Controls (bottom)
p.Val157ArgfsTer115
p.Ala203ArgfsTer74
p.Glu658AlafsTer14
p.Glu872GlyfsTer7
p.Gln102Ter
p.Val157ArgfsTer115
p.Gln166Ter
p.Glu658AlafsTer14
p.Glu854GlyfsTer37
Exons and UTRs
Known mutations
Low complexity domain
Rod domain
Coil
DUF1388

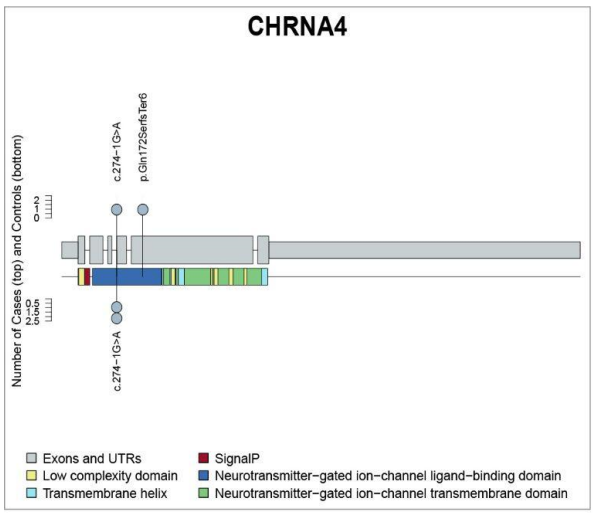

PRPH
Number of Cases (top) and Controls (bottom)
p.Ser15Ter
p.Ala31ProfsTer99
p.Arg64Ter
c.607−1G>A
p.Glu213ThrfsTer6
p.Gly353AlafsTer16
Exons and UTRs
Known mutations
Intermediate filament head, DNA−binding domain
Low complexity domain
Coil
Intermediate filament, rod domain

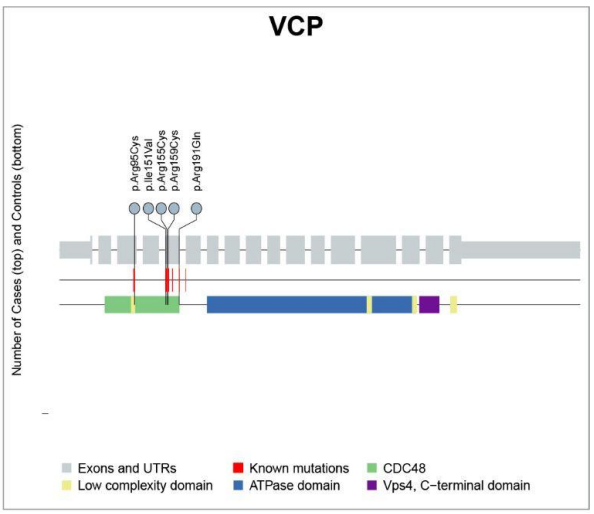

ITPR2
Number of Cases (top) and Controls (bottom)
p.Asn82ThrfsTer8
c.525+1G>A
p.Arg830Ter
c.952−2A>G
p.Arg323Ter
p.Thr643IlefsTer6
c.3679−2A>C
c.4003−2A>G
p.Arg2033Ter
p.His2118GlnfsTer2
p.Gln2552Ter
p.Val1424GlyfsTer15
Exons and UTRs
Coil
IP3R Homology associated domain
IP3−binding region
MIR motif
Transmembrane helix
Low complexity domain
RIH domain
Ion transport domain

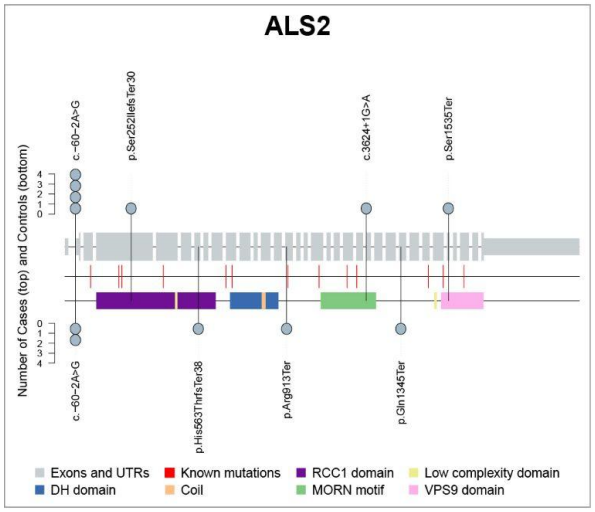

SPG11
Number of Cases (top) and Controls (bottom)
p.Met245ValfsTer2
p.Trp362Ter
p.Asp402IlefsTer14
p.Ile450AsnfsTer26
p.Arg651Ter
p.Leu696ThrfsTer4
p.Asn767LysfsTer4
p.Asn780IlefsTer11
p.Tyr781PhefsTer9
p.Glu1026ArgfsTer4
p.Gln1062Ter
p.Val1072ProfsTer32
p.Cys1107Ter
p.Val1111TrpfsTer6
p.Lys1222IlefsTer15
c.4743+1G>T
c.4906+1G>A
p.Glu1819AlafsTer10
p.Arg1824Ter
p.Glu1921SerfsTer2
p.Leu1950Ter
c.6205+1del
p.Arg2137Ter
p.Arg2286Ter
p.Leu2372Ter
c.7152−1_7156del
p.Glu2406AlafsTer20
p.Met245ValfsTer2
p.Ile450AsnfsTer26
p.Lys1222IlefsTer15
p.Lys1532ArgfsTer21
p.Glu1978Ter
p.Ser2278LeufsTer61
p.Glu2366Ter
Exons and UTRs
Coil
Known mutations
Spatacsin, C−terminal domain
Low complexity domain

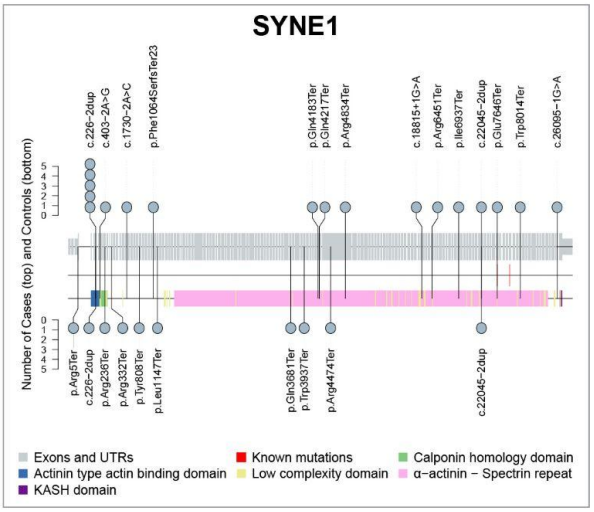

C19orf12
Number of Cases (top) and Controls (bottom)
p.Gly58ArgfsTer10
Exons and UTRs
Transmembrane helix
Low complexity domain

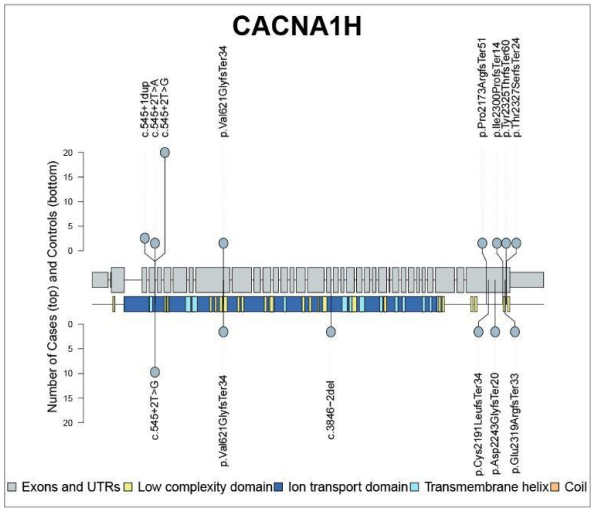

SQSTM1
Number of Cases (top) and Controls (bottom)
c.*855dup
c.*1010_*1014del
c.*1079G>A
c.*1292C>A
p.Tyr140Ter
Exons and UTRs
Low complexity domain
Known mutations
Zinc finger (ZZ type)
PB1 domain
UBA domain

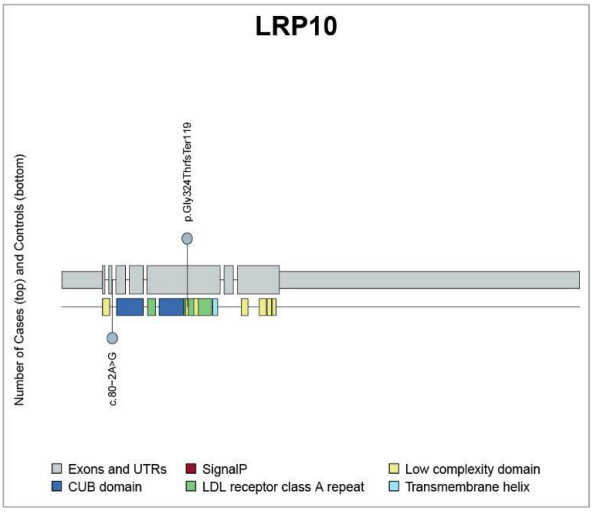

CHRNA4
Number of Cases (top) and Controls (bottom)
c.274−1G>A
p.Gln172SerfsTer6
c.274−1G>A
Exons and UTRs
Low complexity domain
Transmembrane helix
SignalP
Neurotransmitter−gated ion−channel ligand−binding domain
Neurotransmitter−gated ion−channel transmembrane domain

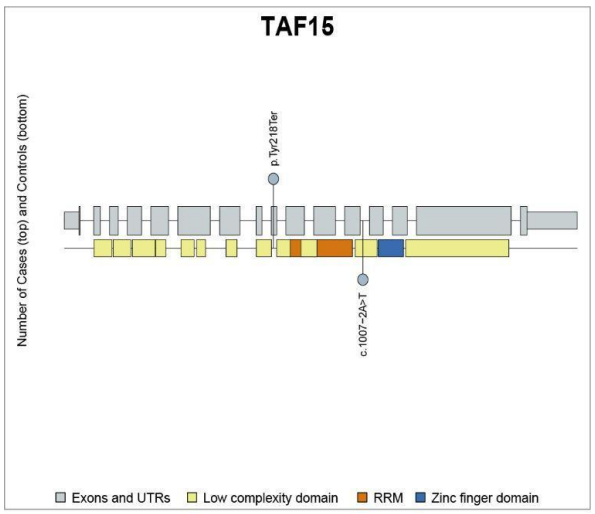

VCP
Number of Cases (top) and Controls (bottom)
p.Arg95Cys
p.Ile151Val
p.Arg155Cys
p.Arg159Cys
p.Arg191Gln
Exons and UTRs
Low complexity domain
Known mutations
ATPase domain
CDC48
Vps4, C−terminal domain

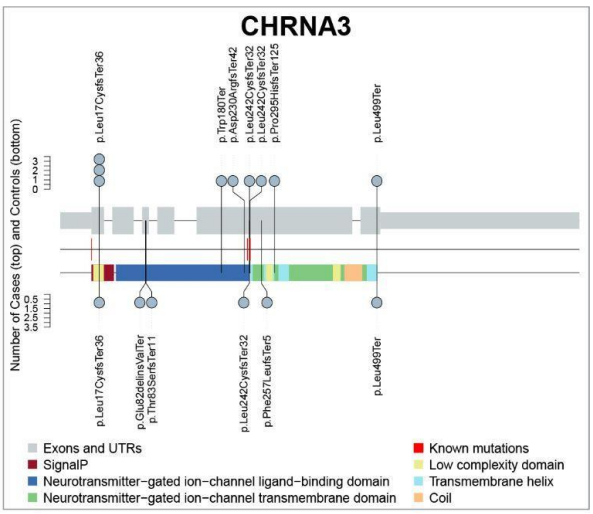

ALS2
Number of Cases (top) and Controls (bottom)
c.−60−2A>G
p.Ser252IlefsTer30
c.3624+1G>A
p.Ser1535Ter
c.−60−2A>G
p.His563ThrfsTer38
p.Arg913Ter
p.Gln1345Ter
Exons and UTRs
DH domain
Known mutations
Coil
RCC1 domain
MORN motif
Low complexity domain
VPS9 domain

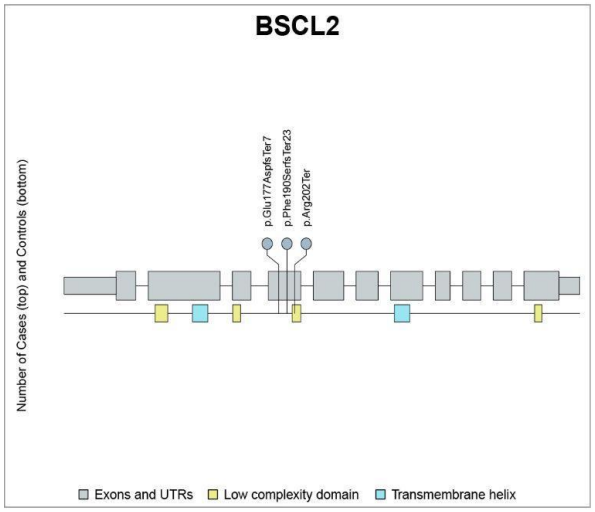

SYNE1
Number of Cases (top) and Controls (bottom)
c.226−2dup
c.403−2A>G
c.1730−2A>C
p.Phe1064SerfsTer23
p.Gln4183Ter
p.Gln4217Ter
p.Arg4834Ter
c.18815+1G>A
p.Arg6451Ter
p.Ile6937Ter
c.22045−2dup
p.Glu7646Ter
p.Trp8014Ter
c.26095−1G>A
p.Arg5Ter
c.226−2dup
p.Arg236Ter
p.Arg332Ter
p.Tyr808Ter
p.Leu1147Ter
p.Gln3681Ter
p.Trp3937Ter
p.Arg4474Ter
c.22045−2dup
Exons and UTRs
Actinin type actin binding domain
KASH domain
Known mutations
Low complexity domain
Calponin homology domain
α−actinin − Spectrin repeat

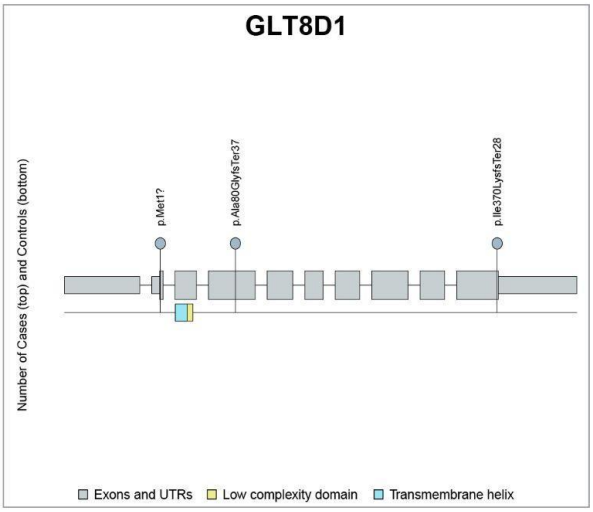

CACNA1H
Number of Cases (top) and Controls (bottom)
c.545+1dup
c.545+2T>A
c.545+2T>G
p.Val621GlyfsTer34
p.Pro2173ArgfsTer51
p.Ile2300ProfsTer14
p.Tyr2325ThrfsTer60
p.Thr2327SerfsTer24
c.545+2T>G
p.Val621GlyfsTer34
c.3846−2del
p.Cys2191LeufsTer34
p.Asp2243GlyfsTer20
p.Glu2319ArgfsTer33
Exons and UTRs
Low complexity domain
Ion transport domain
Transmembrane helix
Coil

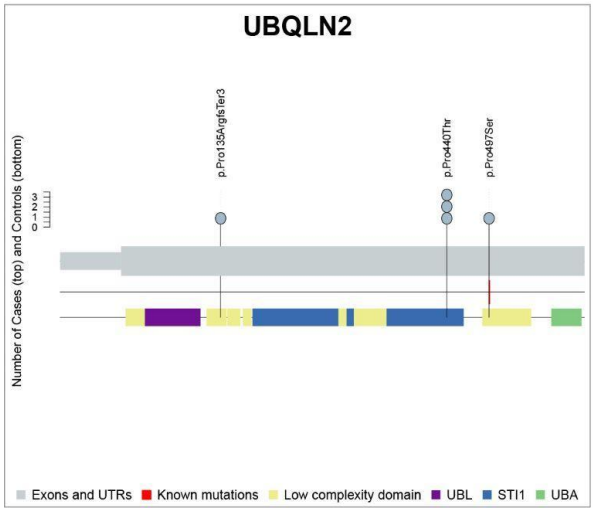

LRP10
Number of Cases (top) and Controls (bottom)
p.Gly324ThrfsTer119
c.80−2A>G
Exons and UTRs
CUB domain
SignalP
LDL receptor class A repeat
Low complexity domain
Transmembrane helix

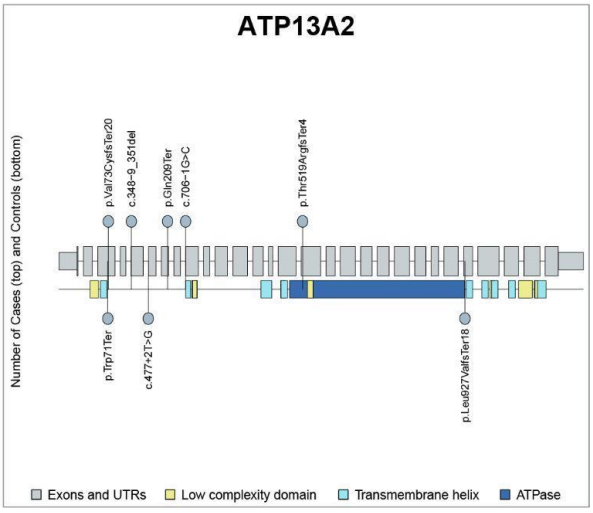

TAF15
Number of Cases (top) and Controls (bottom)
p.Tyr218Ter
c.1007-2A>T
Exons and UTRs
Low complexity domain
RRM
Zinc finger domain

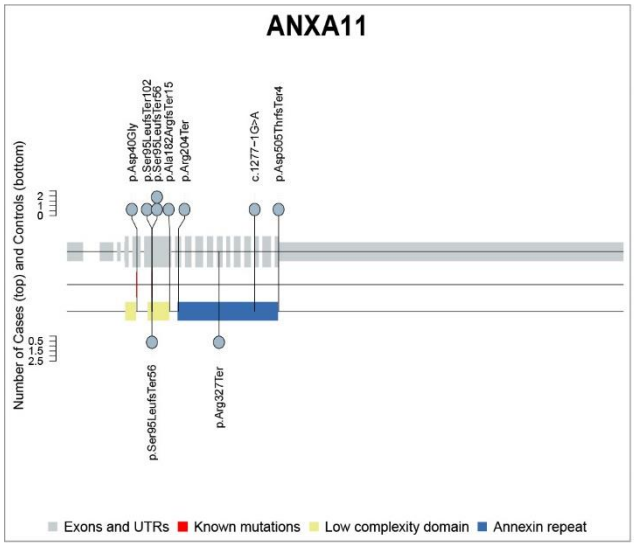

CHRNA3
Number of Cases (top) and Controls (bottom)
p.Leu17CysfsTer36
p.Trp180Ter
p.Asp230ArgfsTer42
p.Leu242CysfsTer32
p.Leu242CysfsTer32
p.Pro295HisfsTer125
p.Leu499Ter
p.Leu17CysfsTer36
p.Glu82delinsValTer
p.Thr83SerfsTer11
p.Leu242CysfsTer32
p.Phe257LeufsTer5
p.Leu499Ter
Exons and UTRs
SignalP
Neurotransmitter-gated ion-channel ligand-binding domain
Neurotransmitter-gated ion-channel transmembrane domain
Known mutations
Low complexity domain
Transmembrane helix
Coil

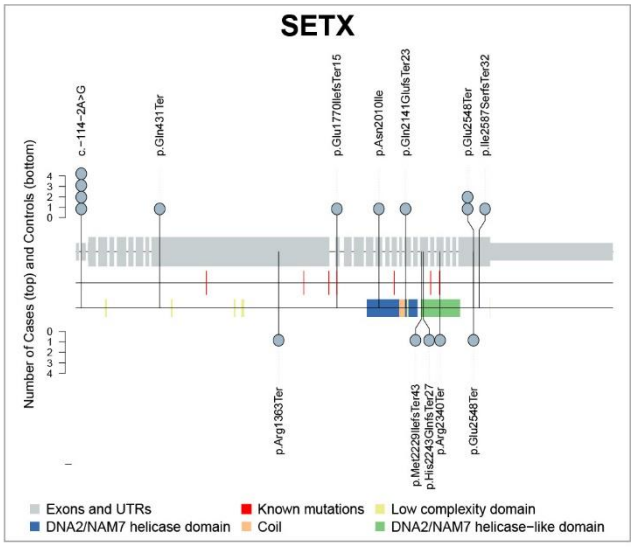

BSCL2
Number of Cases (top) and Controls (bottom)
p.Glu177AspfsTer7
p.Phe190SerfsTer23
p.Arg202Ter
Exons and UTRs
Low complexity domain
Transmembrane helix

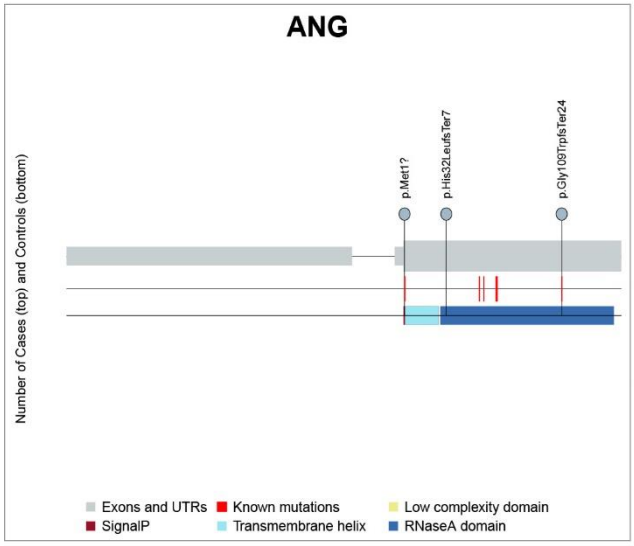

GLT8D1
Number of Cases (top) and Controls (bottom)
p.Met1?
p.Ala80GlyfsTer37
p.Ile370LysfsTer28
Exons and UTRs
Low complexity domain
Transmembrane helix

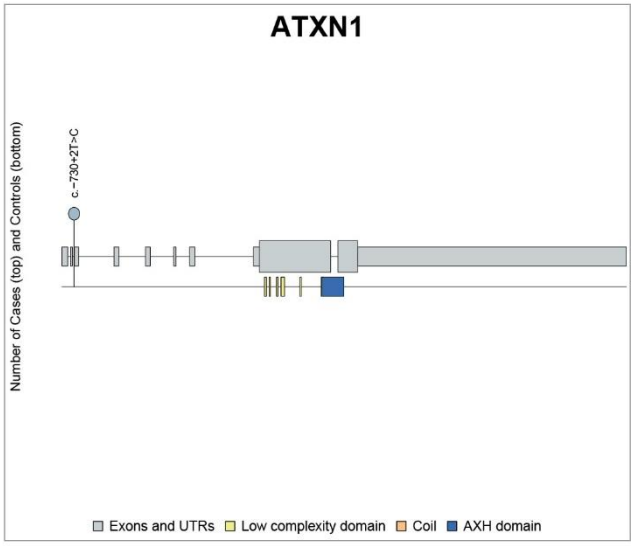

UBQLN2
Number of Cases (top) and Controls (bottom)
p.Pro135AlafsTer3
p.Pro440Thr
p.Pro497Ser
Exons and UTRs
Known mutations
Low complexity domain
UBL
STI1
UBA

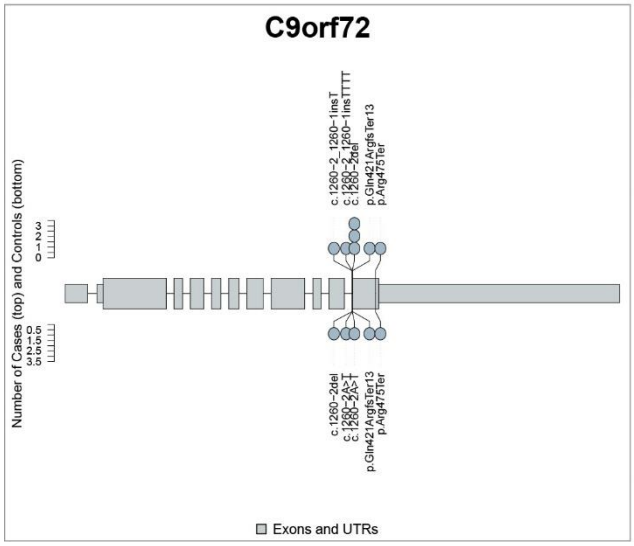

ATP13A2
Number of Cases (top) and Controls (bottom)
p.Val73CysfsTer20
c.348-9_351del
p.Gln209Ter
c.706-1G>C
p.Thr519AlafsTer4
p.Trp71Ter
c.477+2T>G
p.Leu927ValfsTer18
Exons and UTRs
Low complexity domain
Transmembrane helix
ATPase

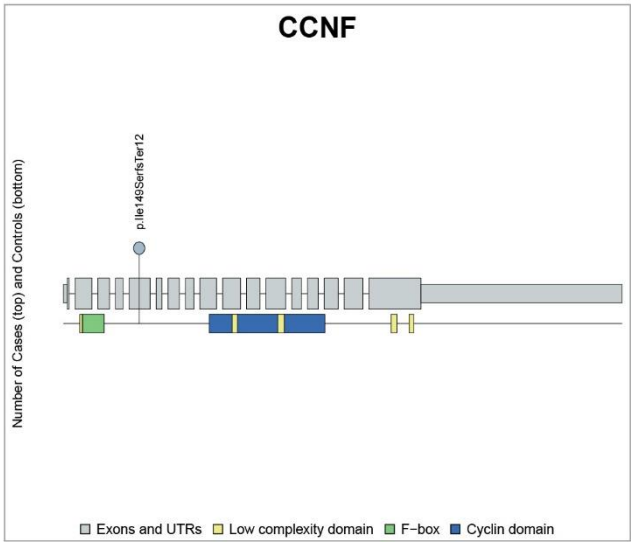

ANXA11
Number of Cases (top) and Controls (bottom)
p.Asp40Gly
p.Ser95LeufsTer102
p.Ser95LeufsTer56
p.Ala182ArgfsTer15
p.Arg204Ter
c.1277−1G>A
p.Asp505ThrfsTer4
p.Ser95LeufsTer56
p.Arg327Ter
Exons and UTRs
Known mutations
Low complexity domain
Annexin repeat

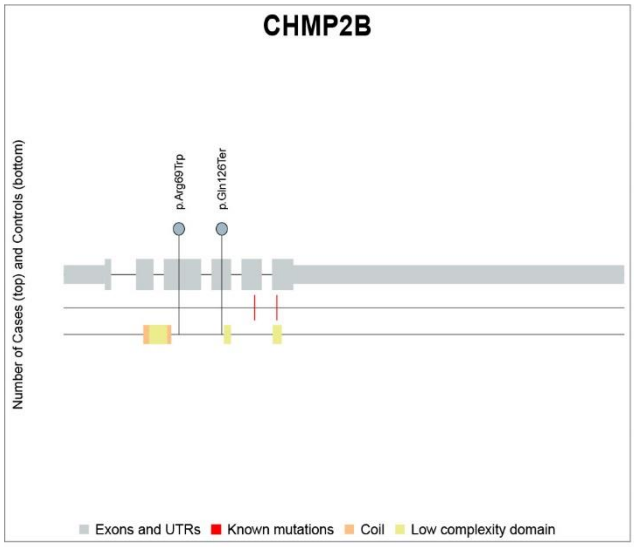

SETX
Number of Cases (top) and Controls (bottom)
c.−114−2A>G
p.Gln431Ter
p.Glu1770IlefsTer15
p.Asn2010Ile
p.Gln2141GlufsTer23
p.Glu2548Ter
p.Ile2587SerfsTer32
p.Arg1363Ter
p.Met2229IlefsTer43
p.His2243GlnfsTer27
p.Arg2340Ter
p.Glu2548Ter
Exons and UTRs
Known mutations
Low complexity domain
DNA2/NAM7 helicase domain
Coil
DNA2/NAM7 helicase−like domain

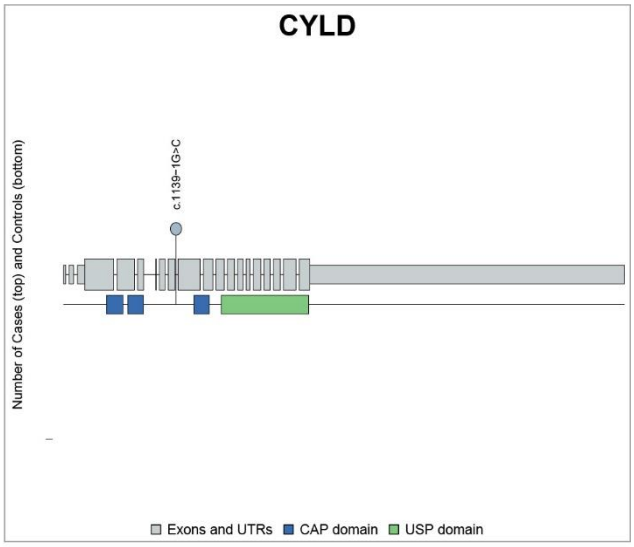

ANG
Number of Cases (top) and Controls (bottom)
p.Met1?
p.His32LeufsTer7
p.Gly109TrpfsTer24
Exons and UTRs
Known mutations
Low complexity domain
SignalP
Transmembrane helix
RNaseA domain

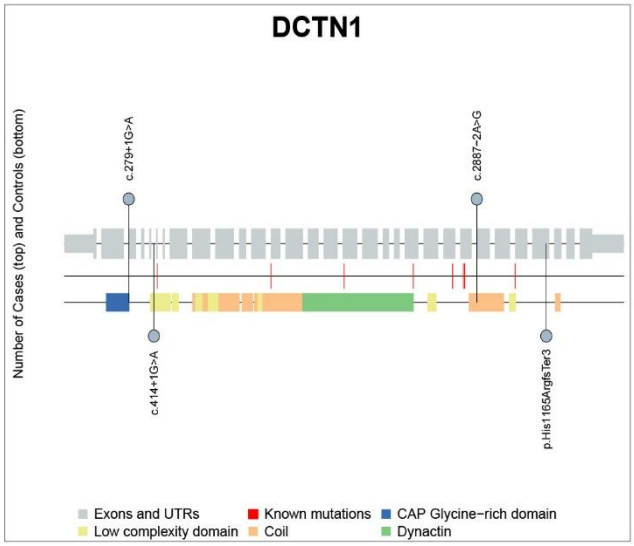

ATXN1
Number of Cases (top) and Controls (bottom)
c.−730+2T>C
Exons and UTRs
Low complexity domain
Coil
AXH domain

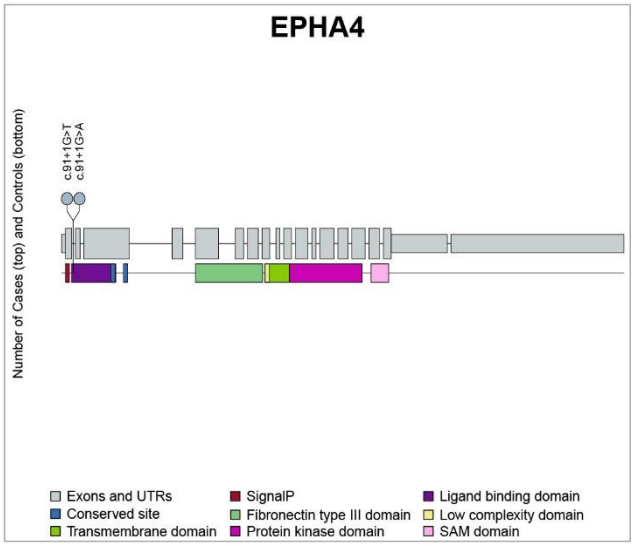

C9orf72
Number of Cases (top) and Controls (bottom)
c.1260−2_1260−1insT
c.1260−2_1260−1insTTTT
c.1260−2del
p.Gln421ArgfsTer13
p.Arg475Ter
Exons and UTRs


CCNF
Number of Cases (top) and Controls (bottom)
p.Ile149SerfsTer12
Exons and UTRs
Low complexity domain
F−box
Cyclin domain


CHMP2B
Number of Cases (top) and Controls (bottom)
p.Arg69Trp
p.Gln126Ter
Exons and UTRs
Known mutations
Coil
Low complexity domain


CYLD
Number of Cases (top) and Controls (bottom)
c.1139−1G>C
Exons and UTRs
CAP domain
USP domain


DCTN1
Number of Cases (top) and Controls (bottom)
c.279+1G>A
c.2887−2A>G
c.414+1G>A
p.His1165ArgfsTer3
Exons and UTRs
Known mutations
CAP Glycine−rich domain
Low complexity domain
Coil
Dynactin


EPHA4
Number of Cases (top) and Controls (bottom)
c.91+1G>T
c.91+1G>A
Exons and UTRs
SignalP
Ligand binding domain
Conserved site
Fibronectin type III domain
Low complexity domain
Transmembrane domain
Protein kinase domain
SAM domain

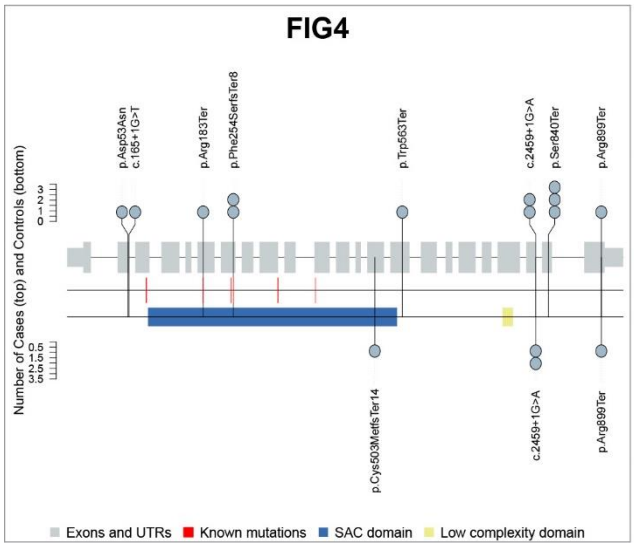
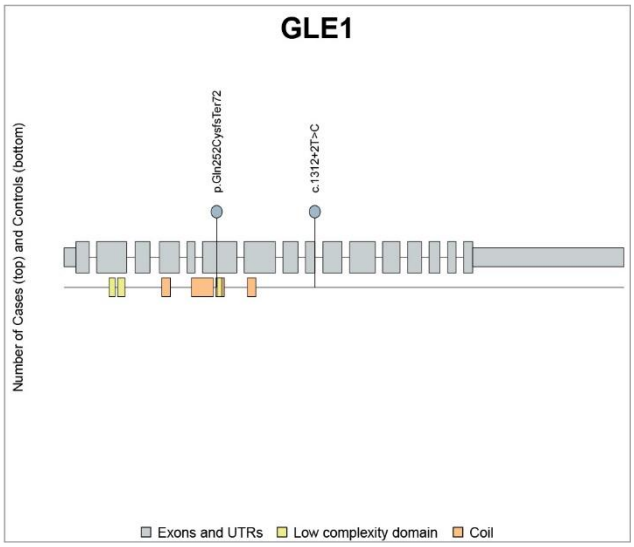
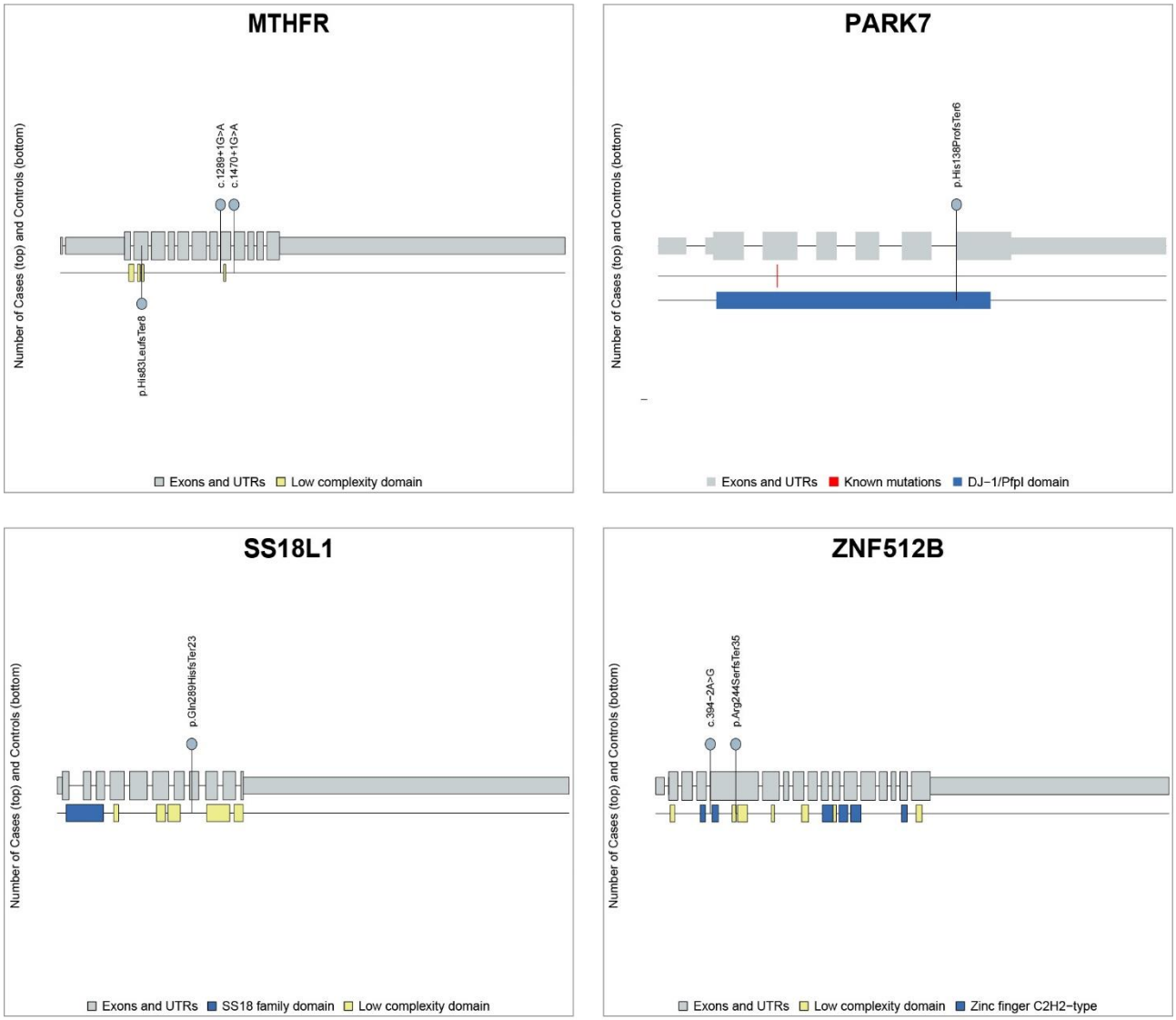


FIG4
Number of Cases (top) and Controls (bottom)
p.Asp53Asn
c.165+1G>T
p.Arg183Ter
p.Phe254SerfsTer8
p.Trp563Ter
c.2459+1G>A
p.Ser840Ter
p.Arg899Ter
p.Cys503MetfsTer14
c.2459+1G>A
p.Arg899Ter
Exons and UTRs
Known mutations
SAC domain
Low complexity domain


GLE1
Number of Cases (top) and Controls (bottom)
p.Gln252CysfsTer72
c.1312+2T>C
Exons and UTRs
Low complexity domain
Coil



Supplementary figure 3: Lolliplots showing pathogenic and likely pathogenic variants encountered in cases and controls.

Lolliplots show the mutational spectra of the displayed genes in cases (n=6013) and controls (n=2411).
For each gene, exons and UTRs are represented as grey bars, connected by introns shown as grey lines (reduced with a factor 1/100). Protein domains and motifs are shown as colored bars. Each lollipop represents a pathogenic or likely pathogenic variant encountered in cases (top) or controls (bottom). The absolute number of cases and controls with the variant is reflected by the height of the lollipop. Known ALS-associated mutations (from ClinVar and the NMHP) are shown as red lines.

Abbreviations: ALS, amyotrophic lateral sclerosis; DH, dbi homology; NMHP, neuromuscular homepage; RCC1, regulator of chromosome condensation 1; RRM, RNA recognition motif domain; SAM, sterile alpha motif; UBA, ubiquitin-associated domain; UBL, ubiquitin-like domain; USP, Ubiquitin specific protease; UTR, untranslated region.

References supplementary data

1. Tweedie S, Braschi B, Gray K, Jones TEM, Seal RL, Yates B, et al. Genenames.org: the HGNC and VGNC resources in 2021. Nucleic Acids Res 2021;49:D939–46.
2. Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD) [Internet] 2021. https://omim.org/.
3. Pestronk A. Neuromuscular homepage-ALS 2021. https://neuromuscular.wustl.edu/.
4. Bonne G, Rivier F, Hamroun D. GeneTable 2021. https://www.musclegenetable.fr/.

5. Abel O, Powell JF, Andersen PM, Al-Chalabi A. ALSoD: A user-friendly online bioinformatics tool for amyotrophic lateral sclerosis genetics. Human Mutation 2012;33:1345–51. https://doi.org/10.1002/humu.22157.
6. GTEx Portal 2021. https://gtexportal.org/home/.
7. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, 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 2015;17:405–24.
8. Jian X, Boerwinkle E, Liu X. In silico prediction of splice-altering single nucleotide variants in the human genome. Nucleic Acids Res 2014;42:13534–44.
9. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res 2018;46:D1062–7.
10. Fadista J, Oskolkov N, Hansson O, Groop L. LoFtool: a gene intolerance score based on loss-of-function variants in 60 706 individuals. Bioinformatics 2017;33:471–4.

Chapter 4: Genetic yield and phenotype-genotype correlations in a large ataxia-HSP cohort in Belgium

Manuscript in preparation.

Abstract

BACKGROUND: Next generation sequencing has revolutionized the diagnostic possibilities for phenotypic and genetic heterogeneous disorders as hereditary ataxia and spastic paraplegia (SPG), with reported yields up to 50%. However, most studies use stringent inclusion criteria, making it difficult to estimate the yield in an unselected population presenting in real-life clinical practice. Also, a considerable number of cases remain unexplained and it is not sure whether Mendelian mutations can close the diagnostic gap.

METHODS: We retrospectively studied a large, real-life cohort of 262 ataxia and 172 SPG patients presenting at the tertiary neuromuscular reference center in University Hospitals Leuven, Leuven, Belgium. We defined the yield of gene-by-gene testing and gene panel analysis. Additionally, we looked for heterozygous variants in genes linked with autosomal recessive ataxia or SPG.

RESULTS: With gene-by-gene testing followed by gene panel analysis, we can explain 54.6% of cases. The yield of gene-by-gene testing was 36.2% and of a subsequent gene panel 28.9%. The latter revealed 46 novel variants, which were entered in the ClinVar database. Friedreich ataxia was the most common diagnosis for ataxia patients, followed by SCA6 and *RFC1*-linked CANVAS. For SPG, the most common subtype was SPG4, followed by SPG7. We identified 31 heterozygous (likely) pathogenic variants in genes linked with autosomal recessive ataxia or SPG, and a combination of variants in eight patients.

CONCLUSIONS: Combined gene-by-gene testing and gene panel analysis provides a high yield for a real-life cohort of ataxia and SPG patients.

Introduction

Hereditary ataxia and spastic paraplegia (SPG) are increasingly seen as the ends of a spectrum, given the clinical, genetic and mechanistic overlap between them.[1] Patients presenting with these phenotypes often pose a diagnostic challenge for clinicians. Ataxia and SPG are manifestations of a wide range of disorders, both acquired and genetic, and clinical presentation is frequently nonspecific with a very broad differential diagnosis.[2,3] Even after thorough diagnostic work-up with imaging, laboratory tests and standard genetic tests, many patients are left without a diagnosis.[4-7] It is assumed that a considerable part of these unexplained patients have a genetic etiology but that (standard) genetic testing has failed to identify the causative genetic defect.[5-8] When testing of the most common ataxia/SPG-associated genes yields negative results, obtaining a genetic diagnosis through gene-by-gene testing is challenging even when a clear familial history is present. Selection of appropriate genes for further testing in an individual patient is difficult given the extensive clinical and genetic heterogeneity and overlap and the lack of discriminating clinical or radiological features pointing to specific subtypes. This makes it often impossible to predict a molecular diagnosis on clinical grounds.[1, 9-11]

The advent of massive parallel sequencing (MPS) has been a major advancement allowing sequencing of large parts of the genome in a relatively fast and cost-effective way. This makes MPS an excellent diagnostic tool for disorders with a long list of partially overlapping disease-causing genes such as ataxia and SPG.[12] Previous sequencing studies have proven the utility of MPS with diagnostic yields between 9.5% and 52% in patients for whom standard tests were negative, even in apparently sporadic patients.[8-10, 13-20] However, use of different inclusion criteria and focus on specific subpopulations (e.g. pediatric versus adult onset, sporadic versus familial cases) in most of these studies makes it difficult to estimate the yield in an unselected population presenting with ataxia/SPG in real-life clinical practice.

In this paper we retrospectively studied a cohort of 262 ataxia and 172 SPG patients presenting at the neurology department of the University Hospital in Leuven, Belgium. We determined the yield of gene-by-gene testing and the added value of gene panel analysis, which we performed for a subset of unexplained patients, and reported novel variants. No patients were excluded based on family history or age of onset, reflecting an unselected cohort consistent with real-life clinical practice. Finally, we looked for heterozygous variants in autosomal recessive (AR) genes as a clue for non-Mendelian inheritance, as we hypothesize that such variants alone may be insufficient to cause disease, but that they might act together with other variants in context of an oligogenic inheritance model.

Methods

Patient selection

Patients were identified by a search in the electronic patient record of the University Hospitals Leuven, Belgium. Medical records of patients presenting at least once at the neurology department between January 2000 and December 2018 (ataxia) and October 2019 (SPG) were screened. Used search terms can be found in *supplementary data*. One of the authors (SVD) reviewed all records.
We included patients with a defined genetic diagnosis as well as patients for whom no etiology was found even after diagnostic work-up. Patients with a clinical diagnosis of multiple system atrophy or other secondary causes of ataxia and/or SPG (e.g. multiple sclerosis) were excluded. Since the study was retrospective and availability of some diagnostic tests changed over time, no strict set of diagnostic investigations was required for inclusion. No patients were excluded based on family history or age of onset.

Data collection

Retrospective chart review was performed. Clinical and genetic data were collected and logged in a standardized way in a database. Clinical data were collected from the first visit at the Neurology department and complemented with data from the second visit provided that this visit took place within 1 year of the initial presentation. To evaluate for each gene the phenotypic involvement of distinct parts of the motor system (cerebellar, pyramidal and peripheral system), we determined for each patient which systems were phenotypically affected. Therefore, data from all visits were used. Definitions used for each inheritance pattern are listed in *supplementary material*.

Genetic testing

Gene-by-gene testing

Genomic DNA was extracted from peripheral leukocytes. Genetic diagnostic testing was performed at the Centre for Human Genetics of the University Hospitals Leuven. Tests that were not routinely performed in Leuven were sent to other reference laboratories. For ataxia, repeat expansion disorders such as spinocerebellar ataxias (SCA), Friedreich's ataxia and in elderly patients fragile X-associated tremor/ataxia syndrome (FXTAS) were excluded first, potentially followed by other single gene tests based on the clinical picture. For SPG, *SPG4* variants were usually the first requested genetic test. However, according to a classic clinical approach, not all tests were performed for each patient.

Whole exome sequencing

Whole exome sequencing (WES) was performed in-house at the Centre for Human Genetics. Patients that were not explained by single gene testing qualified for WES. Since WES was only introduced in our hospital after 2010, not all patients were offered testing. We performed a targeted exome approach, sequencing known disease-associated genes using custom capture kits, followed by targeted analysis of genes included in the ataxia/SPG panel *(supplementary table 1).* More details can be found in *supplementary material*. For the resulting variants, we established a classification of variant pathogenicity based on the American College of Medical Genetics and Genomics (ACMG) – criteria.[21] Variants were reviewed by a multidisciplinary consultation with a geneticist, neurologist and bio-informatician to obtain a consensus opinion for each case. We considered the ACMG Class 4 (likely pathogenic) and Class 5 (pathogenic) variants as relevant to the genetic diagnosis. Putative pathogenic variants were confirmed by Sanger sequencing using custom-designed primers.

Approval of the ethics committee

The Medical Ethics Committee UZ KU Leuven approved the protocol for this study (s61184).
Written informed consent was obtained from all patients for whom next generation sequencing was performed.

Analysis

All data were logged into an SQL relational database (SQL Server 14.0.2027). Statistical analysis was performed with R Studio version 3.6.0. Categorical variables were analyzed with a chi-square test. For continuous variables a student's t-test was used. For multivariate analysis Anova was used. A 5% level of significance was used and statistical tests were two-sided.

Results

Overall results

We included 434 probands with ataxia (n=262) and/or SPG (n=172) in this study. Of the total cohort, 150 patients (34.6%) were explained by gene-by-gene testing. For 128 of the 284 remaining patients, we received informed consent to perform a gene panel analysis obtaining a molecular diagnosis in 37 of them (yield of 28.9% or 8.5% of total patient population). Three patients unexplained after gene panel analysis had a clinical diagnosis of cerebellar ataxia, neuropathy, vestibular areflexia syndrome (CANVAS) and were found to carry biallelic *RFC1* repeat expansions (RE) when this genetic test became available.[22] This brought the total yield of gene-by-gene testing to 153/434 (35.3%).

Demographic and clinical characteristics

The main cohort characteristics are illustrated in Table 1.

There was a male predominance in the cohort, but this was not significant in the total cohort ($p=6.8x10^{-2}$) nor in the ataxia cohort ($p=0.3$) or the SPG cohort ($p=9.3x10^{-2}$) separately.

Globally, the mean age at onset for ataxia patients (40.3±21.7 years (mean±SD); n=249) was higher than the mean age at onset for SPG patients (32.9±20.6 years; n=157; $p=7.1x10^{-4}$).

The mean age at onset had a bimodal distribution for ataxia patients (peak at ±10 and ±54 years old) as well as for SPG patients (peak at ±8 and ±44 years old) (Supplementary Fig. 1). Based on the density distribution, we divided each cohort in a group with a younger age at onset (age <27 years for ataxia and <25 years for SPG) and an older age at onset. The frequencies of molecular diagnoses in each group are described in Supplementary Table 2. The mean age at onset of unexplained patients (ataxia: 43.4±21.3 years; n=150 and SPG: 34.0±21.5 years; n=88) was significantly higher than that of patients for whom a molecular diagnosis was obtained in the ataxia cohort, but not in the SPG cohort (ataxia: 35.5±21.4 years; n=99; $p=4.9x10^{-3}$ and SPG: 31.4±19.6 years; n=69; p=0.4). For genetically explained patients with ataxia, patients with a dominant mutation (47.2±16.1 years; n=42) were significantly older at onset than patients with recessive ataxia (22.3±17.9 years; n=50; $p=5.5x10^{-10}$), even after correction for RE status ($p=5.3x10^{-10}$) (Supplementary Fig. 2). For SPG, no significant difference was found between age at onset of patients with dominant mutations (34.0±19.5 years; n=47) compared to patients with recessive mutations (26.7±19.0 years; n=19; p=0.2). The prevalence of the different inheritance patterns is documented in Table 1. To verify if the predicted inheritance patterns correlated well with the transmission patterns of the actual mutations (if one was identified), we listed the inheritance suggested by pedigree analysis and compared these with the actual transmission patterns of the identified mutations in Table 2.

The diagnostic delay showed a trend towards a shorter diagnostic delay in patients presenting more recently for evaluation, but without corresponding higher diagnostic yields (Supplementary Fig. 3).

	ATAXIA COHORT	SPG COHORT	TOTAL
All patients			
Number of probands	262	172	434
Male	139 *(53.1%)*	97 *(56.4%)*	236 *(54.4%)*
Ancestry			
European	250	158	408
African	2	2	4
Asian	4	4	8
Middle-East	3	2	5
America	0	2	2
Russia	0	1	1
Unknown	3	3	6
Age at onset *(years)*: mean [min-max; n]	40.3 [0-78; n=249]	32.9 [0-81; n=157]	37.4 [0-81; n=406]
Age at presentation *(years)*: mean [min-max; n]	50.6 [17-84; n=262]	45.9 [17-87; n=172]	48.8 [17-87; n=434]
MRI brain performed	227 *(86.6%)*	124 *(72.1%)*	351 *(80.9%)*
MRI spine performed	97 *(37.0%)*	121 *(70.3%)*	218 *(50.2%)*
Lumbar puncture performed	74 *(28.2%)*	53 *(30.8%)*	127 *(29.3%)*
Inheritance pattern			
Suggestive inheritance pattern	126 *(48.1%)*	108 *(62.8%)*	234 *(53.9%)*
AD history	41 *(15.6%)*	54 *(31.4%)*	95 *(21.9%)*
AR history	18 *(6.9%)*	11 *(6.4%)*	29 *(6.7%)*
X-linked history	5 *(1.9%)*	1 *(0.6%)*	6 *(1.4%)*
Mitochondrial history	1 *(0.4%)*	0 *(0%)*	1 *(0.2%)*
Ambiguous/uncertain history	61 *(23.3%)*	43 *(25.0%)*	104 *(24.0%)*
Sporadic	136 *(51.9%)*	63 *(36.6%)*	199 *(45.9%)*
Consanguinity	4 *(1.5%)*	6 *(3.5%)*	10 *(2.3%)*
Phenotypic features			
Ataxia	262 *(100%)*	20 *(11.6%)*	282 *(65.0%)*
Spasticity	102 *(38.9%)*	172 *(100%)*	274 *(63.1%)*
Peripheral neuropathy	82 *(31.3%)*	26 *(15.1%)*	108 *(24.9%)*
Results			
Gene-by-gene testing performed *(% of total)*			
No	3 *(1.1%)*	8 *(4.7%)*	11 *(2.5%)*
Yes	259 *(98.9%)*	164 *(95.3%)*	423 *(97.5%)*
Explained by gene-by-gene testing *(% of total of patients for whom gene-by-gene testing was performed)*			
No	163 *(62.9%)*	107 *(65.2%)*	270 *(63.8%)*
Yes	96 *(37.1%)*	57 *(34.8%)*	153 *(36.2%)*
MSP performed *(% of total of patients unexplained after gene-by-gene testing)*			
No	107 *(65.6%)*	35 *(32.7%)*	142 *(52.6%)*
Yes	56 *(34.4%)*	72 *(67.3%)*	128 *(47.4%)*
Explained by MPS *(% of total of patients for whom MPS was performed)*			
No	28 *(50.0%)*	34 *(47.2%)*	62 *(48.4%)*
No, but variants of interest found	12 *(21.4%)*	17 *(23.6%)*	29 *(22.7%)*
Yes	16 *(28.6%)*	21 *(29.2%)*	37 *(28.9%)*
Unexplained overall *(% of total)*	150 *(57.3%)*	94 *(54.7%)*	244 *(56.2%)*
Explained patients			
Type of mutation			
Repeat expansion	86	0	86
Non-repeat expansion molecular	26	78	104

diagnosis			
Age at onset *(years)*: mean [min-max; n]	35.5 [3-78; n=99]	31.4 [0-69; n=69]	33.8 [0-78; n=168]
Diagnostic delay[a] *(years)*: mean [minimum-maximum; n]	11.6 [-16-43; n=98]	18.1 [0-56; n=66]	14.2 [-16-56; n=164]
Diagnostic delay for patients diagnosed by gene-by-gene testing	9.9 [-16-36; n=82]	16.7 [0-56; n=46]	12.3 [-16-56; n=128]
Diagnostic delay for patients diagnosed by NGS	20.4 [5-43; n=16]	21.5 [4-56; n=20]	21.0 [4-56; n=36]
Diagnostic delay for patients with repeat expansion disorder	8.6 [-16-36; n=73]	NA	8.6 [-16-36; n=73]
Diagnostic delay for patients with non-repeat expansion molecular diagnosis	20.3 [5-43; n=25]	18.1 [0-56; n=66]	18.7 [0-56; n=91]
Unexplained patients			
Age at onset *(years)*: mean [min-max; n]	43.4 [0-76; n=150]	34.1 [0-81; n=88]	39.9 [0-81; n=238]

Table 1: Characteristics of the cohort of ataxia and SPG patients. Percentages between brackets are in function of total number of cases.

[a]Diagnostic delay is time between onset and genetic diagnosis. This can be negative when the patient was as known carrier in follow-up before developing first symptoms.

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; max, maximum; min, minimum; n, total number of cases for whom data were available; NA, not applicable; MPS, massive parallel sequencing; SPG, spastic paraplegia.

Ataxia							
	Sporadic	**Suggestive for AD**	**Suggestive for AR**	**Suggestive for X-linked**	**Suggestive for mitochondrial**	**Ambiguous**	**Total**
AD	10	29	2	0	0	9	50
AR	34	0	12	0	0	9	55
X	0	0	0	5	0	2	7
M	0	0	0	0	0	0	0
Unknown	92	12	4	0	1	41	150
	136	41	18	5	1	61	262
SPG							
AD	8	34	1	0	0	13	56
AR	8	0	5	0	0	6	19
X	3	0	0	0	0	0	3
M	0	0	0	0	0	0	0
Unknown	44	20	5	1	0	24	94
	63	54	11	1	0	43	172

Table 2: Comparison of the inheritance suggested by pedigree analysis with the inheritance of the identified mutation.

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; M, mitochondrial; SPG, spastic paraplegia; X, X-linked.

Gene-by-gene testing

For 11 probands (3 ataxia, 8 SPG), no gene-by-gene testing was registered in our database. Reasons were non-testing because of reimbursement problems or because genetic tests had already been performed in a first or second degree family member with similar symptoms. For SPG patients, additional reasons were: loss to follow-up around the year 2000 (*SPAST* was identified as cause of SPG4 in 1999 so testing was not performed in these patients [23]) or immediate performance of MPS not preceded by gene-by-gene testing. The average number of tests per index

patient for ataxia patients was 3.3 (range 0-11; n=262) and for SPG 2.0 (range 0-8, n=172). The two SCA series offered in our center (first series: SCA 1-2-3-6-7 and second series: SCA 8-10-12-17 + dentatorubral-pallidoluysian atrophy) were considered as one test each. The number of cases for whom SCA series 1 and 2, *FXN* and *SPAST* were tested and the corresponding yields are described in Supplementary table 3.

The total yield of gene-by-gene diagnostic screening was 36,2% (153/423) (37.1% for ataxia and 34.8 % for SPG); patients without gene-by-gene testing excluded. One ataxia patient carried an intermediate repeat of unknown significance in *TBP* (SCA17), which was also present in the unaffected father (additional gene panel showed no pathogenic variants). The distribution of the genetic subtypes identified by gene-by-gene testing is shown in Fig. 1 (A2, B2*)*. The yield per inheritance pattern is shown in Fig. 3A.

Gene panel analysis

A gene panel was requested and performed for 128 patients (56 ataxia and 72 SPG). Demographic features for this subcohort were comparable with the features of the full cohort *(Supplementary table 4)*. For 37 of these patients (28.9%) a molecular diagnosis was found and for 29 additional patients (22.7%) variants of interest were identified (Table 3; Fig. 1 (A3,B3); Fig. 2). 'Variants of interest' were defined as variants that we felt were possibly associated with the disease, but for which we had insufficient genetic/clinical proof of pathogenicity to declare them as genetic cause in a diagnostic setting. Besides, for some recessively transmitted genes as *SPG7*, *STUB1* and *KIF1A*, evidence is emerging that heterozygous variants might be pathogenic as well (possibly with reduced penetrance). Therefore heterozygous variants in these genes were also classified as variant of interest.[24-26] However, the c.728C>T variant in *STUB1* has been described in the heterozygous state in another family and is probably pathogenic.[27] The yield per inheritance pattern is shown in Fig. 3B. In total, we found 20 novel pathogenic variants and 26 additional novel variants of interest. These variants were entered in the ClinVar database.[28] The most commonly mutated genes were *SPG7* (12 patients), *SETX*, *SYNE1*, *CYP7B1 and KIF5A* (three patients each). Most variants of interest were found in *KIF1A*, *REEP1* and *SPG7* (seven, four and three patients respectively). *SYNE1* mutations were first reported to cause pure cerebellar ataxia, but pyramidal and peripheral nerve involvement can be part of the phenotype as well, as is clearly illustrated in our data (Table 3).[29] Likewise, prominent ataxia in patients with *KIF5A* mutations has only recently come under attention, as we found in patient 1247.[30] In Fig. 4 we illustrated for each gene how frequently patients with a mutation in that gene showed phenotypic involvement of the cerebellar, pyramidal and peripheral system.

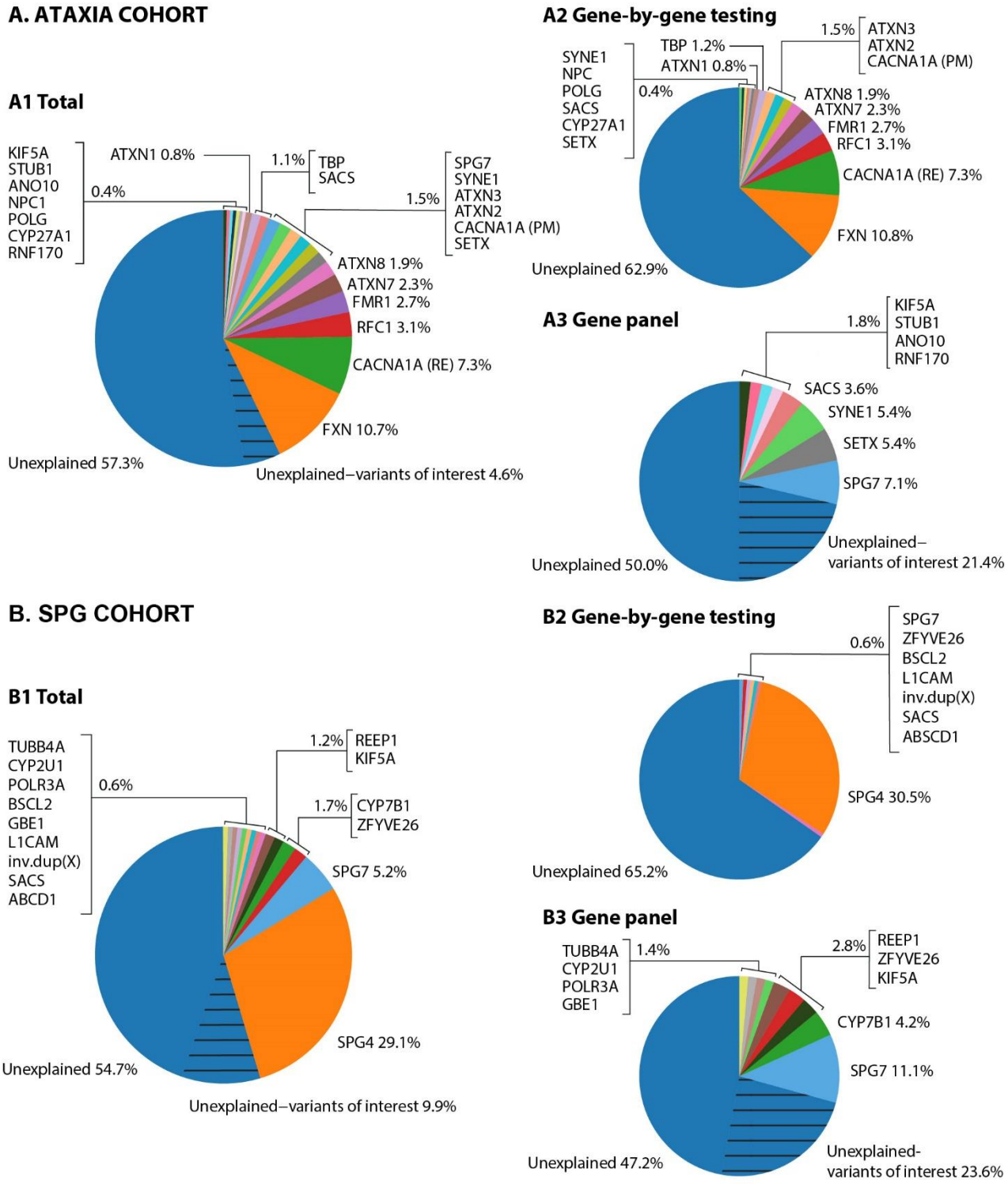


Figure 1: Diagnostic yield and frequency of molecular diagnoses for the total cohort, for gene-by-gene testing and for gene panel analysis, separately for ataxia (A) and SPG (B).

Some genes (e.g. *SYNE1*, *SACS*, *SETX*) were identified both through gene-by-gene testing and by gene panel analysis. The explanation for this is that the availability of gene-by-gene testing for some genes changed over time. A second reason is that some gene-by-gene tests were not requested by the clinician because the phenotype was not suggestive (enough) and it was chosen to directly perform a gene panel analysis.

Abbreviations: inv, inversion; dup, duplication; PM, point mutation; RE, repeat expansion; SPG, spastic paraplegia.

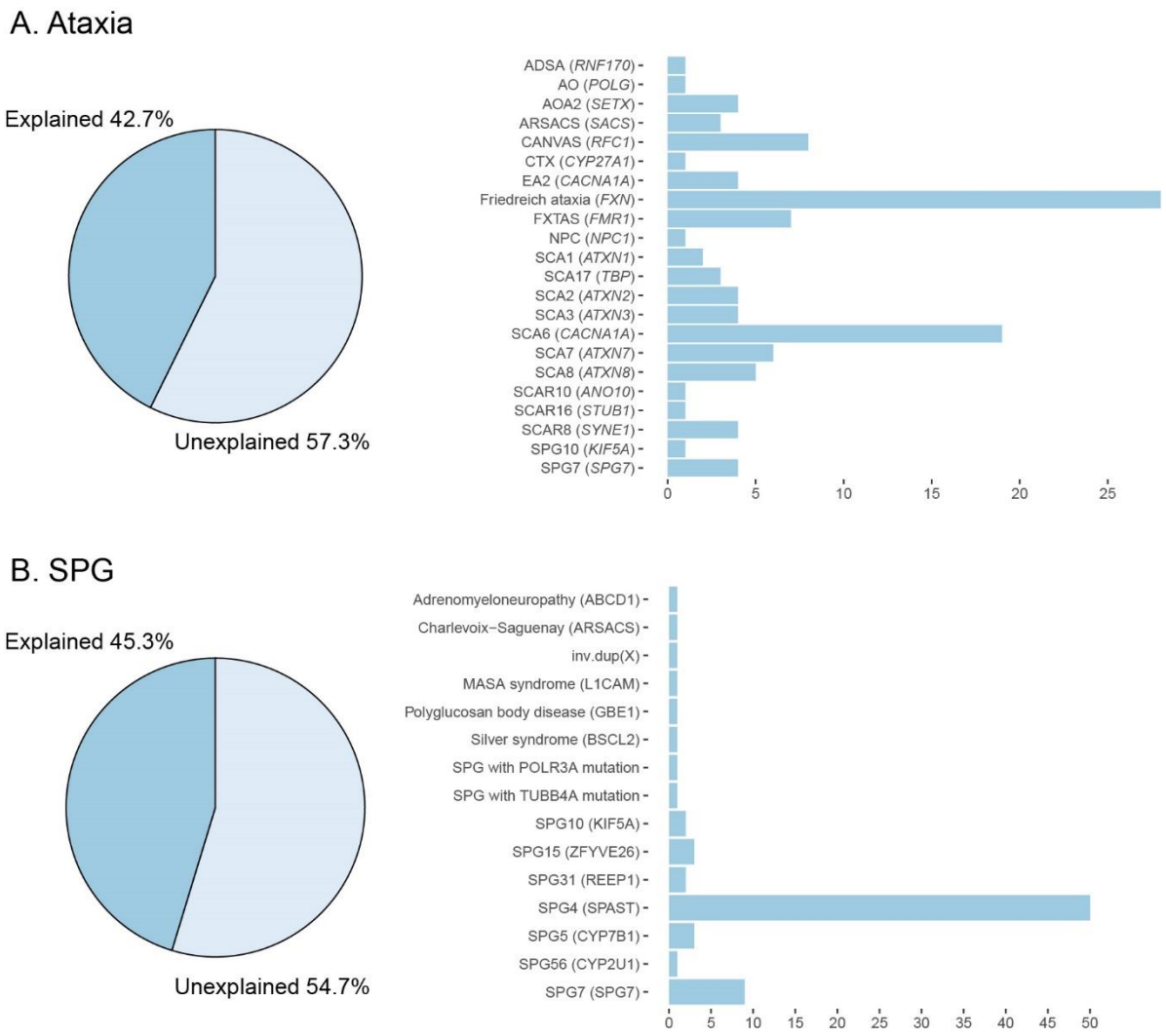


Figure 2: Proportion of patients explained and unexplained as well as absolute number of explained patients with a certain genetic diagnosis for the ataxia (A) and SPG (B) cohort.

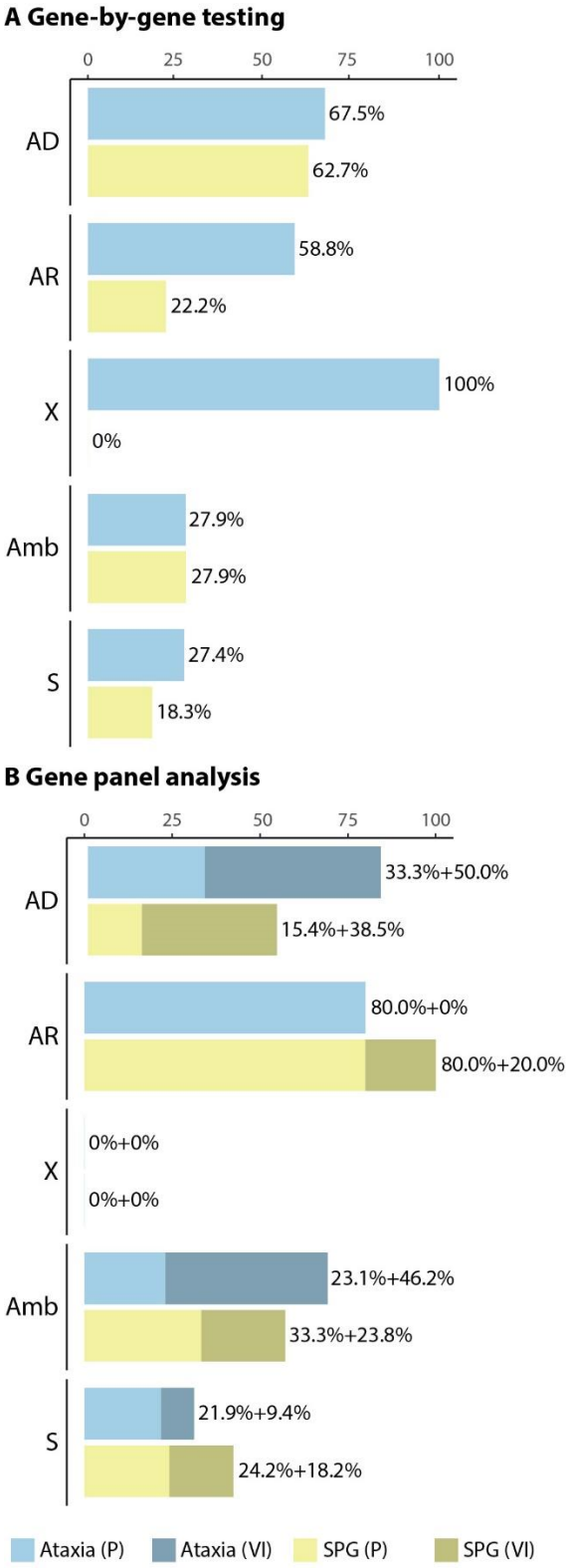


Figure 3: Yield of gene-by-gene testing (A) and gene panel analysis (B) in function of family history.
Results for the ataxia cohort are displayed in blue (light: pathogenic (P); dark: variant of interest (VI)), results for the SPG cohort in yellow (light: pathogenic (P); dark: variant of interest (VI)). The high yield (100%) in ataxia patients with an X-linked inheritance can be explained by the fact that these patients had family members (grandchildren, nephews) already known with the fragile X-syndrome so that the diagnosis of Fragile X-associated tremor ataxia syndrome (FXTAS) was strongly suspected.

Abbreviations : AD, autosomal dominant; Amb, ambiguous; AR, autosomal recessive; P, pathogenic; S, sporadic; SPG, spastic paraplegia; VI, variant of interest; X, X-linked

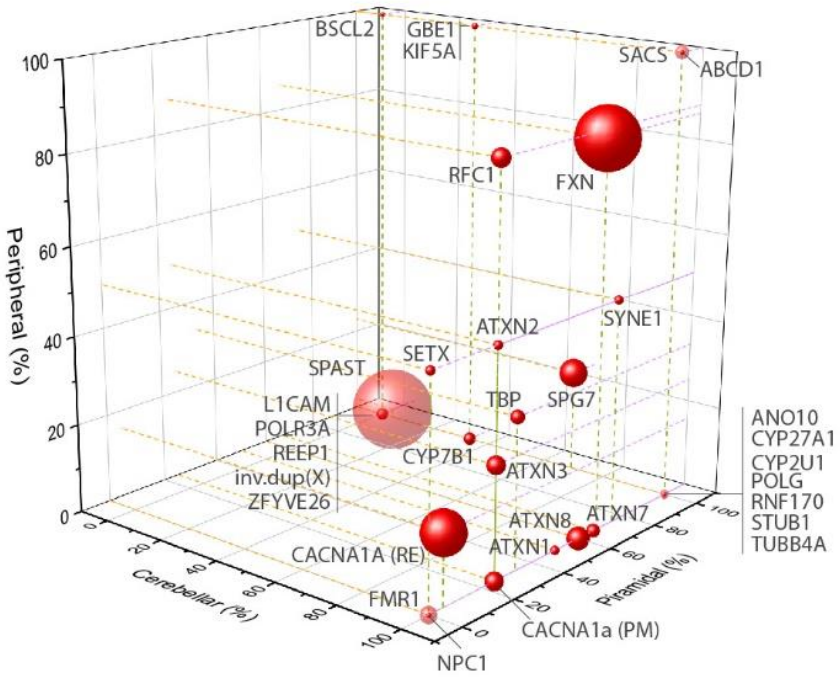


Figure 4: 3D plot illustrating the phenotypic involvement of different parts of the motor system (cerebellar, pyramidal and peripheral system) for each gene identified in our cohort (either by gene-by-gene testing or by gene panel analysis).

Each gene is represented by a globe. The diameter of the globe corresponds to the number of cases with a pathogenic mutation in that gene. For each gene, the proportion of cases with clinical involvement of the cerebellar, peripheral and pyramidal system is represented by the coordinates on the x-, y- and z-axis respectively.

Abbreviations: PM, point mutation; RE, repeat expansion.

Heterozygous variants in genes linked with autosomal recessive disorders

Since heterozygous variants in *SPG7* were enriched in a Canadian SPG cohort, we scrutinized our cohort for heterozygous variants in genes linked with AR ataxia or SPG (Table 4).[31] In our MPS data, we found 12 patients carrying 13 heterozygous (likely) pathogenic variants. No second hits were found in the respective genes. Additionally, 12 heterozygous variants were yet listed as variants of interest in Table 3, because mutations might sometimes exert a dominant effect in these genes as well, leading to similar phenotypes (three in *SPG7*, two in *STUB1* and six in *KIF1A*), or because we found biochemical abnormalities (mildly reduced sterole levels and a variant in *CYP7B1*) *(*Table 4*)*. Lastly, gene-by-gene testing had shown four patients with a heterozygous RE in *FXN* (for three of them point mutations on the other allele were excluded).

One patient carried three heterozygous variants, eight other patients carried a combination of a heterozygous variant with a pathogenic variant or a variant of interest.

ID, Sex	Co-hort	AO	IP	CB signs	Py signs	Pe signs	Cognitive impairment	Other	MRI brain/ spine	Gene	Variants	Repor-ted in litera-ture?
	EXPLAINED											
1052, F	AT	27	AR	LA, GA, SE	HR	Mild re-innervation on EMG with normal NCS in LL	N	/	Cerebellar atrophy	*ANO10*	c.269_270delAA; p.Lys90Argfs*4[a] c.1477-2A>G[a]	N N
1070, M	AT	23	S	LA, GA, D, SE, OD	HR, HT	N	N	/	Cerebellar atrophy	*SPG7*	c.1529C>T; p.Ala510Val[a] c.861dupT; p.Asn288*fs*1[a]	Y[44] Y[44]
1117, M	AT	3	S	LA, GA, D, Ny	HT, B	Mixed SM neuropathy	Y	/	Cerebellar atrophy	*SACS*	HOM [c.12270_12273del; p.His4090Glnfs*10]	N
1122, F	AT	27	Amb	LA, GA, D, SE, OD, Ny	B	Axonal SM neuropathy	Subjective memory impairment[b]	Migraine Dystonia	Cerebellar atrophy	*SETX*	c.6322C>T; p.Gln2108*[a] c.718+3G>C [a,c]	Y[45] N
1162, F	AT	13	S	LA, D, SE[b], Ny	N	Axonal SM neuropathy	N	/	Cerebellar atrophy	*SETX*	c.6322C>T; p.Gln2108* c.6638C>T; p.Pro2213Leu	Y[45] Y[46]
1182, F	AT	12	Amb	LA, GA, D, Ny	HT, B	Mixed SM neuropathy	N	/	Cerebellar hypoplasia, LH pons	*SACS*	HOM [c.8869_8870insTTT;8871del]	N
1201, M	AT	26	S	LA, GA, D, SE	HR, HT	N	Y	/	Cerebellar atrophy	*STUB1*	c.688_690delTAC; p.Tyr230del[a] c.722G>A; p.Arg241Gln[a]	N N
1203, V	AT	37	Amb	LA, GA, D, OD	HR, HT	N	N	Sarcoidosis	Cerebellar atrophy	*SYNE1*	c.19603G>T; p.Glu6535* c.23767_23768del[a]; p.Gln7923Glufs*4[a]	N Y[47
1206, M	AT	42	S	LA, GA, D, OD	HR, HT	Axonal sensory neuropathy	N	/	Cerebellar atrophy	*SYNE1*	c.23020-1G>A c.23767_23768del; p.Gln7923Glufs*4	N Y[47]
1220, M	AT	19	S	LA, GA, Ny	HR, HT	N	N	Visual complaints	Cyst between optic chiasm and hypo-thalamus	*SPG7*	c.2228T>C; p.Ile743Thr c.1045G>A; p.Gly349Ser	Y[48] Y[44,48]
1239, M	AT	30	AR	LA, SE, SS	HR, B	N	N	Vertigo	Cerebellar atrophy	*SPG7*	c.1529C>T; p.Ala510Val[a] c.861dupT; p.Asn288*fs*1[a]	Y[44] Y[44]

1241, M	AT	13	AR	LA, GA, SE, SS, OA	N	NA	N	Diplopia	N	*SETX*	HOM [c.5308_5311del; p.Glu1770Ilefs*15]	Y[49]
1242, F[d]	AT	35	AD	LA, GA, SE	HT, B	HoR but no neuropathy on NCS/EMG	N	/	Cortico-subcortal atrophy	*RNF170*	c.595C>T; p.R199C[a]	Y[50]
1247, M	AT	47	AD	LA, GA, D, SE, SS	HR, HT, B	Mixed SM neuropathy	Y	/	Arachnoid cyst	*KIF5A*	c.2744_2754del + c.2743T>G; p.Ser915Cysfs*24 + p.Ser915Ala	N
1261, M	AT	32	AR	LA, GA, D, OD	HR, HT	Axonal SM neuropathy	N	/	Cerebellar atrophy	*SYNE1*	c.3130C>T; p.Arg1044* c.23767_23768del; p.Gln7923Glufs*4	N Y[47]
2353, M	AT	15	S	LA, D	HR, HT	Mild calf atrophy, chronic re-innervation on EMG with normal NCS	N	Migraine Mild extra-pyramidal signs	Cerebellar atrophy	*SPG7*	HOM [c.1529C>T; p.Ala510Val]	Y[44]
1188, F	SPG	11	Amb	LA, GA	Spastic gait, B	Axonal motor neuropathy	N	Visual complaints[b]	N	*SPG7*	HOM [c.1454_1462del; p.Arg485_Glu487del]	Y[48]
2367, F	SPG	34	Amb	N	HR, HT, B	N	N	/	N	*REEP1*	c.183-2A>T[a]	N
2393, F	SPG	27	AD	N	HR, HT, B	Axonal SM neuropathy	N	/	N	*KIF5A*	c.611G>A; p.Arg204Gln	Y[51]
2394, M	SPG	53	Amb	N	HR, HT	Mixed SM neuropathy	N	Ptosis	N	*SPG7*	c.861dupT; p.Asn288*fs*1 c.1529C>T; p.Ala510Val	Y[44] Y[44]
2406, M	SPG	NA	AD	N	HR, HT, B	Myopathic changes on EMG	N	/	N	*REEP1*	c.224T>A; p.Leu75Gln[a]	N
2421, F	SPG	15	Amb	N	HT, B	HoR, no neuropathy on NCS/EMG	N	Scoliosis	N	*POLR3A*	c.2323_2329del, p.Asn775ProfsTer19 c.1909+22G>A	N Y[52]
2425, F	SPG	49	Amb	N	HR, HT, B	N	N	/	N	*SPG7*	c.861dupT; p.Asn288*fs*1 c.1529C>T; p.Ala510Val	Y[44] Y[44]
2439, M	SPG	1	AR	N	HR, HT, B	N	Y	/	N	*ZFYVE26*	HOM [c.2756-6G>A][c]	N

2440, M	SPG	38	AR	N	Spastic gait, B	N	N	/	NA	*CYP7B1*	c.1456C>T; p.Arg486Cys c.825T>A; p.Tyr275*	Y[53] Y[53]
2442, M	SPG	43	S	LA	HR, HT, B	N	N	/	N	*SPG7*	c.1454_1462del; p.Arg485_Glu487del c.1529C>T; p.Ala510Val	Y[54] Y[44]
2468, F	SPG	26	AR	N	HR, HT, B	N	N	/	Hyper-intensity GP	*SPG7*	HOM [c.1729G>A; p.Gly577Ser]	Y[55]
2471, M	SPG	36	S	LA	HR, HT, B	N	N	/	N	*SPG7*	HOM [c.233T>A; p.Leu78*]	Y[32]
2480, M	SPG	52	S	N	HT, B	Axonal SM neuropathy	N	Deficient glycogen branching enzyme in fibroblasts	WML	*GBE1*	c.2081T>A; p.Ile694Asn c.760A>G; p.Thr254Ala	N Y[56]
2481, M	SPG	40	S	N	HR, HT	N	N	/	Mild cerebellar atrophy	*SPG7*	c.1045G>A; p.Gly349Ser c.1053dupC[a]; p.Gly352Argfs*44[a]	Y[44,48] Y[54]
2489, M	SPG	13	S	N	HR, HT, B	NA	N	ADD	N	*CYP7B1*	HOM [c.825T>A; p.Tyr275*][a]	Y[57]
2499, F	SPG	27	S	N	HR, HT, B	Axonal motor neuropathy	N	ADD	N	*KIF5A*	c.610C>T; p.Arg204Trp[a]	Y[57]
2501, M	SPG	2	Amb	LA, Ny, SS	HT, B	NA	Y	Optic neuropathy AVS	WML	*TUBB4A*	c.286G>A, p.Gly96Arg[a]	Y[58]
2506, M	SPG	49	Amb	LA, GA	HR, HT, B	N	N	/	WML	*CYP7B1*	c.1456C>T; p.Arg486Cys[a] c.1155C>G; p.Tyr385*[a]	Y[57] N
2512, M	SPG	1	AR	Ny, SS	HR, HT, B	NA	Y	RT	NA (calci-fication GP on CT)	*CYP2U1*	HOM [c.471del; p.Ile158Serfs*2][a]	N
2520, MM	SPG	14	S	N	HR, HT	Axonal motor neuropathy	N	/	Peri-ventricular WML	*ZFYVE26*	HOM [c.2214G>A; p.Trp738*][a]	N
2533, M	SPG	10	S	LA	HT	Axonal motor neuropathy	N	/	N	*SPG7*	c.1045G>A; p.Gly349Ser c.2140A>C; p.Thr714Pro	Y[44,48] N

	VARIANTS OF INTEREST											
1048, F	AT	48	AD	LA, GA, D, Ny	B[b]	N	N	Headache	Cerebellar atrophy	*KIF1A*	c.1999T>A; p.Tyr667Asn	N
1072, M	AT	42	Amb	LA	HR[b], HT[b]	N	N	Epilepsy Episode of chorea	Cerebellar atrophy Myelo-malacia on MRI spine	*STUB1*	c.728C>T; p.Pro243Leu	Y[27,59]
1076, F	AT	35	AD	LA, GA, D, SE	HR, HT	N	Y	Tremor Epilepsy Extra-pyramidal signs	Cerebellar atrophy	*STUB1*	c.688_690del; p.Tyr230del	N
1160, M	AT	2	Amb	LA, D, Ny	HR	N	N	Vertigo	Transient T2 hyper-intensities cerebellum	*MME*	c.27dupT; p.Ile10Tyrfs*3[e] c.855+6A>T (LB)	N N
1164, M	AT	6	AD	LA	HT	HoR, no neuropathy on NCS/EMG	Low IQ	Concentric con-striction of visual fields	Venous anomaly in cerebellum	*CACNA1A* *ITPR1*	c.4997G>A; p.Arg1666His[f] c.2420G>A; p.Ser807Asn[f]	Y[9] N
1165, F	AT	53	Amb	LA, SE, Ny, adducti-on deficit right eye	HR, HT	Axonal SM neuropathy	Low IQ	Epilepsia at younger age	Cerebellar and brainstem atrophy	*KIF1A*	c.1031C>T; p.Thr344Met c.799-4G>A	N N
1193, M	AT	45	S	LA, GA	N	Axonal SM neuropathy	Y	Epilepsia	Cerebellar atrophy	*TMEM240* *GRM1*	c.325G>C; p.Gly109Arg c.1602+3A>G	N N
1197, M	AT	10	S	LA, GA, Ny	HT	N	N	/	Cerebellar atrophy	*SACS* *POLG*	c.12590A>G; p.Tyr4197Cys[a] c.8393C>A; p.Pro2798Gln[a] c.1760C>T, p.Pro587Leu[g] c.752C>T, p.Thr251Ile[g]	N Y[60] Y[61] Y[61]
1226, F	AT	48	Amb	LA[b], GA[b]	HR, HT	N	N	/	Cerebellar atrophy	*PRKCG*	c.206G>T; p.Cys69Phe	N
1227, M	AT	42	S	LA, GA	N	chronic re-innervation	N	Epilepsia	N	*WASHC5*	c.2426C>G; p.Pro809Arg	N

						on EMG with normal NCS						
1234, F	AT	43	Amb	LA, GA	N	N	N	/	N	*CYP7B1*	c.825T>A; p.Tyr275*[h]	Y[57]
1260[i], F	AT	39	S	LA, GA	N	Axonal sensory neuropathy	Subjective memory impairment	Cough Weight loss	Cerebellar atrophy	*SPG7*	c.1053del, p.Gly352Alafs*87	N
2336, M	AT	6	Amb	LA, D	HR, HT	Axonal SM neuropathy	N	/	Cerebellar atrophy	*CAPN1*	HOM [c.1441C>G; p.Arg481Gly]	N
2363, M	SPG	32	S	N	HR, HT	Axonal SM neuropathy and myopathic changes on EMG	N	/	Aspecific WML in left frontal region	*REEP1*	c.136G>A; p.Ala46Thr	N
2364, F	SPG	44	AD	N	HR, HT, B	N	N	/	Restricted periventricular WML	*TUBB4A*	c.784C>T, p.Arg262Cys	N
2376, M	SPG	39	S	N	HR, HT	N	N	/	Venous angioma and pineal cyst	*SPG7*	c.1442C>G; p.Thr481Arg	N
2424, F	SPG	NA	Amb	N	HR, B	N	N	Migraine	N	*REEP1*	c.145A>C; p.Thr49Pro	N
2461, M	SPG	NA	Amb	N	HR, HT, B	NA	N	/	N	*KIF1A*	c.756C>G; p.Ser252Arg[j]	Y[26]
2469, F	SPG	20	Amb	LA	HR, HT, B	N	N	/	N	*SPG7*	c.1454_1462del; p.Arg485_Glu487del	Y[34]
2476, F	SPG	50	AD	N	HR, HT	N	N	/	N	*KIF1A*	c.1103T>C; p.Leu368Pro[k]	N
2482, M	SPG	56	S	N	HR, HT	N	N	/	NA	*BSCL2*	c.630G>A; p.Ser210Ser	N
2485, M	SPG	36	S	N	HR, HT	N	N	/	N	*KIF1A*	c.790C>G; p.Arg264Gly[l]	N
2487, M	SPG	0	AD	N	HT	Axonal motor neuropathy	N	/	NA	*ATL1*	c.1346C>A; p.Ala449Asp	N

2490, M	SPG	38	S	N	HR, HT	NA	N	/	Posterior WML and atrophy	*WASHC5*	c.2422A>G; p.Ile808Val	Y[62]
2491, M	SPG	34	AR	N	HT, B	DM neuropathy	N	/	N	*KIF1A*	c.4639G>A; p.Glu1547Lys	N
2510, M	SPG	4	S	N	HT, B	N	N	/	N	*KIF1A*	c.3498_3500del; p.Glu1166_Tyr1167delinsAsp[m]	N
2511, M	SPG	1	AD	N	HR, HT, B	Increased insertional activity	N	MI	N	*REEP1*	c.145A>C ; p.Thr49Pro[a]	N
2529, F	SPG	71	AD	SE	HR, HT, B	N	N	/	Choroidal fissure cyst	*REEP1*	c.145A>C ; p.Thr49Pro[a]	N
2530, F	SPG	10	Amb	LA, GA, D	HR, HT	NA	N	/	WML	*CAPN1*	c.1126C>T + c.1796G>C; p.Arg376Trp + p.Arg599Pro[a] c.2145A>G; p.*715Trpext*103[a]	N N
2537, F	SPG	53	Amb	N	HR, HT	N	N	/	N	*KCND3*	c.1004T>G; p.Phe335Cys[a]	N

Table 3: Patients with pathogenic variants or variants of interest identified by gene panel analysis: phenotypic characteristics and identified variants.
Cerebellar and pyramidal signs described are from the first visit only, unless otherwise stated. Results from MRI and nerve conduction studies (NCS)/electromyography (EMG) and 'other' signs can result from later dates.
Reference sequences used are: NM_0187075.5 (*ANO10*); NM_001127713.1 (*ATL1*); NM_001122955.3 *(BSCL2)*; NM_023035.3 (*CACNA1A*); NM_001198868.2 (*CAPN1*); NM_183075.3 *(CYP2U1)*; NM_004820.5 (*CYP7B1*); NM_000158.4 (*GBE1*); NM_001278064.2 (*GRM1*); NM_001168272.2 (*ITPR1*); NM_004980.5(*KCND3*); NM_001244008.2 (*KIF1A*); NM_004984.4 (*KIF5A*); NM_007288.3 (*MME*); NM_001126131.2 (*POLG*); NM_007055.4 *(POLR3A)*; NM_002739.5 (*PRKCG*); NM_022912.3(*REEP1*); NM_030954.4 (*RNF170*); NM_014363.6 (*SACS*); NM_015046.7 (*SETX*); NM_003119.4 (*SPG7*); NM_005861.4 (*STUB1*); NM_182961.4 (*SYNE1*); NM_001114748.2 (*TMEM240*); NM_006087.4 (*TUBB4A*); NM_014846.4 (*WASHC5*); NM_015346.4 (*ZFYVE26*)

[a]Segregation analysis performed: the variant segregated with disease in the family members available
[b]Sign developed after the first visit
[c]RNA analysis showed an effect on splicing
[d]This patient has been described before[50]
[e]Mother of this patients is homozygous for this variant, which is probably the cause of her axonal neuropathy. The implication of the heterozygous variant in the patient is unknown. Other, currently unaffected family members are heterozygous for the variant.
[f]Variant present in affected father as well
[g]Both mutations have been described in combination before. In these studies, both mutations were found on the same allele and were compound heterozygous with another pathogenic mutation. Segregation analysis for these variants has not been performed for our patient.[61]
[h]Biochemical analyses showed an atypical sterole profile with mildly elevated levels of 27-hydroxycholesterol. No second variant in *CYP7B1* was found.
[i]This patient has another molecular diagnosis that explains her symptoms: CANVAS (homozygous *RFC1* repeat expansions)

[j]Segregation analysis was performed but the results are ambiguous. The affected brother carried the same variant. Patient's mother also carried the variant, but it was unclear whether she was affected or not. Due to practical reasons, the mother could not be evaluated in our hospital.
[k]Segregation analysis was performed but the results are ambiguous. The affected mother and niece carried the same variant. Another niece did not carry the variant, but her clinical status was unsure. This niece has so far declined further clinical evaluation.
[l]Segregation analysis performed: the father also carried the variant, but his clinical status was unsure and he declined further clinical evaluation.
[m]Segregation analysis was declined by the patient and/or family

Abbreviations: Amb, Ambiguous; AD; autosomal dominant; ADD, attention deficit disorder; AO, age at onset; AR, autosomal recessive; AT, ataxia; AVS, aortic valve stenosis; B, babinski's sign; CANVAS, cerebellar ataxia-neuropathy-vestibular areflexia syndrome; CB, cerebellar; D, dysarthria; DM, demyelinating; FDI, first dorsal interosseous; GA, gait ataxia; GP, globus pallidus; HA, hereditary ataxia; HOM, homozygous; HoR, hyporeflexia; HR, hyperreflexia; HT, hypertonia; IP, inheritance pattern; LA, limb ataxia; LB, likely benign; LH, linear hypointensities; MI, mitral insufficiency; N, no; NA, not available; Ny, nystagmus; OA, oculomotor apraxia; OD, ocular dysmetria; Pe, peripheral; Py, pyramidal; RT, Retinal telangiectasia; S, sporadic; SE, saccadic eye movements; SM, sensorimotor; SPG, spastic paraplegia; SS, slow saccades; WML, white matter lesions; Y, yes.

ID, Sex	Phenotype	Gene	Variant	Molecular Diagnosis
MPS: heterozygous (likely) pathogenic variants				
1147, M	Ataxia	*WFS1*	c.1958G>T; p.Arg653Leu	None
1201, M	Ataxia	*ATM*	c.3403-2A>T	*STUB1*
1247, M	Ataxia	*ATM*	c.7875_7876delTGinsGC; p.Asp2625_Ala2626delinsGluPro	*KIF5A*
2420, M	SPG	*WFS1*	c.1294C>G; p.Leu432Val	None
2464, F	SPG	*MTTP*	c.399dupA; p.Glu134Argfs*8	None
2474, M	SPG	*PIK3R5*	c.2034+2T>C	None
2489, M	SPG	*GJC2*	c.89T>C; p.Val30Ala	*CYP7B1*
2490, M	SPG	*CLN5*	c.812A>G; p.Asn271Ser	None (variant of interest in *WASHC5*)
2499, F	SPG	*WWOX*	c.109C>G; p.His37Asp	*KIF5A*
2510, M	SPG	*PEX7*	c.875T>A; p.Leu292*	None (heterozygous variant of interest in *KIF1A*; see below)
		ATM	c.1A>G; p.Met1Val	None (heterozygous variant of interest in *KIF1A*; see below)
2522, F	SPG	*PHYH*	c.135-2A>G	None
2529, F	SPG	*AP4M1*	c.975-1G>T	None (variant of interest in *REEP1*)
MPS: heterozygous variants of interest (also described in table 3)				
1048, F	Ataxia	*KIF1A*	c.1999T>A; p.Tyr667Asn	None
1072, M	Ataxia	*STUB1*	c.728C>T; p.Pro243Leu	None
1076, F	Ataxia	*STUB1*	c.688_690del; p.Tyr230del	None
1234, F	Ataxia	*CYP7B1*	c.825T>A; p.Tyr275*	None
1260, F	Ataxia	*SPG7*	c.1053del, p.Gly352Alafs*87	*RFC1*
2376, M	SPG	*SPG7*	c.1442C>G; p.Thr481Arg	None
2461, M	SPG	*KIF1A*	c.756C>G; p.Ser252Arg	None
2469, F	SPG	*SPG7*	c.1454_1462del; p.Arg485_Glu487del	None
2476, F	SPG	*KIF1A*	c.1103T>C; p.Leu368Pro	None
2485, M	SPG	*KIF1A*	c.790C>G; p.Arg264Gly	None
2491, M	SPG	*KIF1A*	c.4639G>A; p.Glu1547Lys	None
2510, M	SPG	*KIF1A*	c.3498_3500del; p.Glu1166_Tyr1167delinsAsp	None (2 other heterozygous variants, see above)
Gene-by-gene testing: heterozygous variants				
1112, M	Ataxia	*FXN*	Repeat expansion	*RFC1*
1152, F	Ataxia	*FXN*	Repeat expansion	None
1156,F	Ataxia	*FXN*	Repeat expansion	None
1168, M	Ataxia	*FXN*	Repeat expansion	None

Table 4: Heterozygous P and LP variants in genes linked with autosomal recessive ataxia and SPG.
Reference sequences: NM_004722.4 (*AP4M1*); NM_000051.4 (*ATM*); NM_006493.4 *(CLN5)*; NM_004820.5 (*CYP7B1*); NM_020435.4 (*GJC2*); NM_001244008.2 (*KIF1A*); NM_000253.4 (*MTTP*); NM_000288.4 (*PEX7*); NM_006241.4 (*PHYH*); NM_014308.3 (*PIK3R5*); NM_001126131.2 (*POLG*); NM_003119.4 (*SPG7*); NM_005861.4 (*STUB1*); NM_006005.3 (*WFS1*); NM_016373.4 (*WWOX*).

Abbreviations: LP, likely pathogenic; MPS, massive parallel sequencing; P, pathogenic.

Discussion

In this paper, we retrospectively characterized a cohort of ataxia and SPG patients consulting in our tertiary neuromuscular reference center in Leuven, Belgium. Unlike many other studies performing MPS for ataxia and/or SPG, we did not use exclusion criteria based on age of onset or family history to ensure that our cohort reflected real-life clinical practice.[9,14,15,17,20]
The yield of gene-by-gene testing was 36,2% and that of subsequent MPS 28,9%. Together they can provide a genetic diagnosis for 54.6% of patients (68.5% when variants of interest are included). Limitations of our study include the retrospective design with no strict criteria for the exclusion of acquired causes of ataxia and SPG, the fact that we did not perform MPS for all unexplained patients and that we could not compare the frequency of variants to a large local control database.

Epidemiology and phenotype-genotype correlations

Our cohort consisted of 262 ataxia and 172 SPG patients. Epidemiological characteristics as sex ratio, mean age of onset and inheritance mode where comparable with other cohorts described before.[4,7,11]

Overall, approximately half of the cases had a possible positive family history. The inheritance patterns derived from history taking were good predictors for the true inheritance of the mutation if one was identified. However, since three patients with AR inheritance carried AD mutations, we must remain careful in letting it define our diagnostic (genetic) quest and offer testing for genes with alternative inheritances.

Ataxia and SPG are increasingly seen as the extremes of a spectrum and often lumped together in studies, given the clinical, genetic and pathophysiological overlap between them.[6,14,32] However, phenotype-genotype correlations for individual genes remain important for the correct interpretation of variants found in MPS data. We evaluated for each gene the fraction of patients with a mutation in that gene that exhibited cerebellar, pyramidal and/or peripheral signs (Fig. 4). Although numbers are too small to draw major conclusions at individual gene level, many genes affect more than one system, often with a preference for one. On the other hand, heterogeneity is present. This illustrates that although the specific mutation and corresponding gene are strong drivers of pathology, it is not sufficient to explain the full phenotype in an individual patient.[33]

Gene-by-gene testing

Approximately one third of cases was provided with a genetic diagnosis by gene-by-gene testing. This yield is somewhat higher than the yield in a similar Canadian ataxia cohort (26,7%) but in this study only adult onset cases were included and only repeat expansions were tested for.[4]

Although MPS is on the rise to replace gene-by-gene testing as a first line genetic test, the latter still has an important value. Especially repeat expansions are currently not reliably identified by MPS and should thus be requested separately for ataxia patients. For SPG, more than one in four patients screened for *SPAST* mutations were positive. Although *SPAST* mutations are easily identified by MPS, it might be a sensible approach to first screen *SPAST* as a single gene test before performing MPS analysis, depending on the local cost and turnaround time of MPS.
As expected, the yield of gene-by-gene testing was higher in patients with a clear familial history.

Gene panel analysis

The yield of gene panel analysis was 28.9% (28.6% for ataxia and 29.2% for SPG). Since analysis was performed in a diagnostic setting, only convincing variants were reported as pathogenic. However, including 'variants of interest' (since we believe that at least some of them are actually pathogenic), the yield increased to 51.6% (50.0% for ataxia and 52.8% for SPG). This yield is roughly comparable to what has been described by other studies performing MPS for ataxia/SPG patients, but comparison is difficult due to different inclusion criteria, variable prior genetic testing (e.g. *SPG4*) and different genes included in the panel.[9,14,17,19] Most pathogenic variants were identified in genes associated with AR disorders, probably because most prevalent AD disorders were already excluded by gene-by-gene testing (SCAs and SPG4).[34] The yield of gene panel analysis is lower than that of gene by gene testing, but this is partially explained because the patients for whom gene panel analysis was performed were not naïve (this is gene by gene testing had already been performed). For now, given that structural variants such as repeat expansions are not yet reliably identified from (short read) sequencing studies, the two remain complementary. However, in the future it might be possible to detect all these variants from one single technique (e.g. long read sequencing).

The high yield of gene panel analysis highlights the importance of MPS in disorders with extensive genetic and phenotypic heterogeneity as ataxia and SPG, where the yield of gene-by-gene tests declines steeply once the most common subtypes are excluded. When we extrapolate our results, assuming that for all patients gene-by-gene testing, and if negative, a gene panel was performed; 54.6% of the cohort would have received a genetic diagnosis; 68.5% with variants of interest included. The fact that MPS was only performed for 128 of the 270 patients unexplained after gene-by-gene testing could have created a bias. However, most patients for whom no MPS was performed, were not offered because they were lost to follow up before MPS became available. Only a very small proportion of patients were offered MPS but declined *(data not shown)*.
For more than 30%, no genetic diagnosis could be established. Possible reasons are true negatives, structural or noncoding variants, variants in genes not yet associated with the phenotype and not

included in the panel or epigenetic changes. The finding of pathogenic RE in the new gene *RFC1* years after a negative gene panel analysis in three patients illustrates the importance of a regular panel update and reanalysis when new genes are discovered. However, Ngo et al recently illustrated that reanalysis of their exome cohort did not contribute much to the missing heredity and that the maximum detecting rate currently hovers at approximately 50%.[34]

Contemplating non-Mendelian inheritance

An alternative explanation for this diagnostic gap is that a mix of Mendelian and complex/oligogenic inheritance underlies the genetic architecture of ataxia and SPG, as is seen in other neurodegenerative disorders as amyotrophic lateral sclerosis, Parkinson's and Alzheimer's disease.[35-37] Recent findings in *SPG7* might provide a hint to such genetic constitution where not all variants carry the same pathogenic weight. For example, recessive as well as dominant mutations in *SPG7* have been reported, some with variable penetrance and expressivity, although for some cases a second deep intronic variant might have been missed in some cases.[38,39] Hypomorphic mutations (e.g. p.Ala510Val), frequently encountered in the general population and probably (more) pathogenic in trans with certain deleterious variants, have been described.[16,38] Additionally, a recent study found that heterozygous pathogenic *SPG7* variants were enriched in SPG patients, sometimes in combination with an additional pathogenic variant in another SPG gene, suggesting digenic inheritance.[31] Finally, penetrance of CAG repeat expansions linked with hereditary ataxia at the population level has shown to be lower than what is expected based on clinical cohorts.[40] This suggests that other, yet unknown (genetic) factors play a role in the expression of the disease. In our data, we looked for likely pathogenic or pathogenic heterozygous variants in genes associated with AR ataxia and SPG, and found 29 variants in 27 patients. One patient carried three such variants, eight other patients carried a combination of a heterozygous variant with a pathogenic variant or variant of interest. Moreover, the carrier frequency of *FXN* RE in non-Friedreich ataxia patients who were screened (4/127) was higher than expected in controls (1/60-1/90).[41] However, caution is warranted because not all patients received a gene panel and we did not compare with a control population. Further research is needed to confirm that such (combinations of) variants are enriched in comparison with control populations.

Genetic landscape of ataxia and SPG in Belgium

In our ataxia cohort, the most common genetic subtypes were repeat expansions disorders. Friedreich ataxia is the most frequent hereditary ataxia and was with 10.7% also most prevalent in our cohort, all patients carrying homozygous RE.[42] SCA6 was the second most common ataxia subtype. Although SCA3 is regarded as the most prevalent SCA worldwide, we only found 4 patients (1.5%) with *ATXN3* RE.[6] Recently identified *RFC1* RE (CANVAS) explained 3.1% of our ataxia

patients, these patients have been more extensively described elsewhere.[43] The other genetic subtypes were all relatively rare, underscoring the regional variability in genetic causes of ataxia.

In SPG patients, the most common genetic subtype worldwide is SPG4, which was also the case in our cohort (29.1%).[6] The second most frequent subtype was SPG7 (5.2%). SPG11 is often reported as the most prevalent AR SPG, but was absent in our cohort.[6] Similar to the genetic heterogeneity in ataxia, the other genetic subtypes were each relatively rare.

Conclusion

Combining gene-by-gene testing and gene panel analysis in a real-life ataxia-SPG cohort can provide a genetic diagnosis for 54.6% (68.5% when variants of interest are included). Given that repeat expansions are currently not yet reliably distilled from short read massive parallel sequencing data (and especially exome sequencing), gene-by-gene testing remains essential for these phenotypes, currently with a yield of 36.2%. A genetic characterization of ataxia and SPG patients is essential for trial readiness since gene-specific therapies are on the way. The diagnostic gap might be in part explained by non-Mendelian inheritance but further evidence is needed to support this hypothesis.

References

1. Synofzik, M. & Schule, R. Overcoming the divide between ataxias and spastic paraplegias: Shared phenotypes, genes, and pathways. Mov Disord 32, 332-345 (2017).
2. Barsottini, O.G., Albuquerque, M.V., Braga-Neto, P. & Pedroso, J.L. Adult onset sporadic ataxias: a diagnostic challenge. Arq Neuropsiquiatr 72, 232-240 (2014).
3. Tesson, C., Koht, J. & Stevanin, G. Delving into the complexity of hereditary spastic paraplegias: how unexpected phenotypes and inheritance modes are revolutionizing their nosology. Hum Genet 134, 511-538 (2015).
4. Kraft, S., et al. Adult onset spinocerebellar ataxia in a Canadian movement disorders clinic. Can J Neurol Sci 32, 450-458 (2005).
5. Kerber, K.A., Jen, J.C., Perlman, S. & Baloh, R.W. Late-onset pure cerebellar ataxia: differentiating those with and without identifiable mutations. J Neurol Sci 238, 41-45 (2005).
6. Ruano, L., Melo, C., Silva, M.C. & Coutinho, P. The global epidemiology of hereditary ataxia and spastic paraplegia: a systematic review of prevalence studies. Neuroepidemiology 42, 174-183 (2014).
7. Schule, R., et al. Hereditary spastic paraplegia: Clinicogenetic lessons from 608 patients. Ann Neurol 79, 646-658 (2016).
8. Keogh, M.J., et al. Frequency of rare recessive mutations in unexplained late onset cerebellar ataxia. J Neurol 262, 1822-1827 (2015).
9. Coutelier, M., et al. Efficacy of Exome-Targeted Capture Sequencing to Detect Mutations in Known Cerebellar Ataxia Genes. JAMA Neurol 75, 591-599 (2018).

10. 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).
11. Schols, L., et al. Genetic background of apparently idiopathic sporadic cerebellar ataxia. Hum Genet 107, 132-137 (2000).
12. Bamshad, M.J., et al. Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet 12, 745-755 (2011).
13. Nemeth, A.H., et al. Next generation sequencing for molecular diagnosis of neurological disorders using ataxias as a model. Brain 136, 3106-3118 (2013).
14. 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).
15. Sawyer, S.L., et al. Exome sequencing as a diagnostic tool for pediatric-onset ataxia. Hum Mutat 35, 45-49 (2014).
16. Coutelier, M., et al. A panel study on patients with dominant cerebellar ataxia highlights the frequency of channelopathies. Brain 140, 1579-1594 (2017).
17. Burguez, D., et al. Clinical and molecular characterization of hereditary spastic paraplegias: A next-generation sequencing panel approach. J Neurol Sci 383, 18-25 (2017).
18. Kara, E., et al. Genetic and phenotypic characterization of complex hereditary spastic paraplegia. Brain 139, 1904-1918 (2016).
19. Kumar, K.R., et al. Targeted next generation sequencing in SPAST-negative hereditary spastic paraplegia. J Neurol 260, 2516-2522 (2013).
20. Nibbeling, E.A.R., et al. Exome sequencing and network analysis identifies shared mechanisms underlying spinocerebellar ataxia. Brain 140, 2860-2878 (2017).
21. 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, 405-424 (2015).
22. Cortese, A., et al. Biallelic expansion of an intronic repeat in RFC1 is a common cause of late-onset ataxia. Nat Genet 51, 649-658 (2019).
23. Hazan, J., et al. Spastin, a new AAA protein, is altered in the most frequent form of autosomal dominant spastic paraplegia. Nat Genet 23, 296-303 (1999).
24. Sanchez-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).
25. Genis, D., et al. Heterozygous STUB1 mutation causes familial ataxia with cognitive affective syndrome (SCA48). Neurology 91, e1988-e1998 (2018).
26. Pennings, M., et al. KIF1A variants are a frequent cause of autosomal dominant hereditary spastic paraplegia. Eur J Hum Genet 28, 40-49 (2020).
27. Roux, T., et al. Clinical, neuropathological, and genetic characterization of STUB1 variants in cerebellar ataxias: a frequent cause of predominant cognitive impairment. Genet Med 22, 1851-1862 (2020).
28. Landrum, M.J., et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res 46, D1062-D1067 (2018).
29. 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).
30. Qiu, Y., et al. A novel KIF5A gene variant causes spastic paraplegia and cerebellar ataxia. Ann Clin Transl Neurol 5, 1415-1420 (2018).
31. Estiar, M.A., et al. Evidence for Non-Mendelian Inheritance in Spastic Paraplegia 7. Mov Disord 36, 1664-1675 (2021).
32. Pfeffer, G., et al. SPG7 mutations are a common cause of undiagnosed ataxia. Neurology 84, 1174-1176 (2015).
33. Darios, F., Coarelli, G. & Durr, A. Genetics in hereditary spastic paraplegias: Essential but not enough. Curr Opin Neurobiol 72, 8-14 (2021).
34. Ngo, K.J., et al. A diagnostic ceiling for exome sequencing in cerebellar ataxia and related neurological disorders. Hum Mutat 41, 487-501 (2020).

35. van Blitterswijk, M., et al. Evidence for an oligogenic basis of amyotrophic lateral sclerosis. Hum Mol Genet 21, 3776-3784 (2012).
36. Klein, C., Lohmann-Hedrich, K., Rogaeva, E., Schlossmacher, M.G. & Lang, A.E. Deciphering the role of heterozygous mutations in genes associated with parkinsonism. Lancet Neurol 6, 652-662 (2007).
37. Van Cauwenberghe, C., Van Broeckhoven, C. & Sleegers, K. The genetic landscape of Alzheimer disease: clinical implications and perspectives. Genet Med 18, 421-430 (2016).
38. McDermott, C.J., et al. Paraplegin gene analysis in hereditary spastic paraparesis (HSP) pedigrees in northeast England. Neurology 56, 467-471 (2001).
39. Verdura, E., et al. A deep intronic splice variant advises reexamination of presumably dominant SPG7 Cases. Ann Clin Transl Neurol 7, 105-111 (2020).
40. Gardiner, S.L., et al. Prevalence of Carriers of Intermediate and Pathological Polyglutamine Disease-Associated Alleles Among Large Population-Based Cohorts. JAMA Neurol 76, 650-656 (2019).
41. Epplen, C., et al. Differential stability of the (GAA)n tract in the Friedreich ataxia (STM7) gene. Hum Genet 99, 834-836 (1997).
42. Burk, K. Friedreich Ataxia: current status and future prospects. Cerebellum Ataxias 4, 4 (2017).
43. Van Daele, S.H., et al. Diagnostic yield of testing for RFC1 repeat expansions in patients with unexplained adult-onset cerebellar ataxia. J Neurol Neurosurg Psychiatry 91, 1233-1234 (2020).
44. Klebe, S., et al. Spastic paraplegia gene 7 in patients with spasticity and/or optic neuropathy. Brain 135, 2980-2993 (2012).
45. Gonzaga-Jauregui, C., et al. Exome Sequence Analysis Suggests that Genetic Burden Contributes to Phenotypic Variability and Complex Neuropathy. Cell Rep 12, 1169-1183 (2015).
46. Moreira, M.C., et al. Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2. Nat Genet 36, 225-227 (2004).
47. Mademan, I., et al. Multisystemic SYNE1 ataxia: confirming the high frequency and extending the mutational and phenotypic spectrum. Brain 139, e46 (2016).
48. van Gassen, K.L., et al. Genotype-phenotype correlations in spastic paraplegia type 7: a study in a large Dutch cohort. Brain 135, 2994-3004 (2012).
49. Anheim, M., et al. Ataxia with oculomotor apraxia type 2: clinical, biological and genotype/phenotype correlation study of a cohort of 90 patients. Brain 132, 2688-2698 (2009).
50. Van Daele, S.H., et al. RNF170 mutation causes autosomal dominant sensory ataxia with variable pyramidal involvement. Eur J Neurol (2021).
51. Goizet, C., et al. Complicated forms of autosomal dominant hereditary spastic paraplegia are frequent in SPG10. Hum Mutat 30, E376-385 (2009).
52. Minnerop, M., et al. Hypomorphic mutations in POLR3A are a frequent cause of sporadic and recessive spastic ataxia. Brain 140, 1561-1578 (2017).
53. Goizet, C., et al. CYP7B1 mutations in pure and complex forms of hereditary spastic paraplegia type 5. Brain 132, 1589-1600 (2009).
54. 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).
55. Wilkinson, P.A., et al. A clinical, genetic and biochemical study of SPG7 mutations in hereditary spastic paraplegia. Brain 127, 973-980 (2004).
56. Said, S.M., Murphree, M.I., Mounajjed, T., El-Youssef, M. & Zhang, L. A novel GBE1 gene variant in a child with glycogen storage disease type IV. Hum Pathol 54, 152-156 (2016).
57. Morais, S., et al. 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 (2017).
58. Lu, Y., Ondo, Y., Shimojima, K., Osaka, H. & Yamamoto, T. A novel TUBB4A mutation G96R identified in a patient with hypomyelinating leukodystrophy onset beyond adolescence. Hum Genome Var 4, 17035 (2017).

59. 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, 31 (2017).
60. Baets, J., et al. Mutations in SACS cause atypical and late-onset forms of ARSACS. Neurology 75, 1181-1188 (2010).
61. Taanman, J.W., et al. Analysis of mutant DNA polymerase gamma in patients with mitochondrial DNA depletion. Hum Mutat 30, 248-254 (2009).
62. Chrestian, N., et al. Clinical and genetic study of hereditary spastic paraplegia in Canada. Neurol Genet 3, e122 (2017).

Supplementary data

Supplementary material

1 Search terms used in medical records of patients

spinocerebellar ataxia, SCA, spinocerebellar atrophy, spinocerebellar degeneration, hereditary ataxia, Machado-Joseph disease, autosomal dominant cerebellar ataxia, ADCA, olivopontocerebellar atrophy, OPCA, SCAR, ARCA, recessive ataxia, Charlevoix-Saguenay, Friedreich, ataxia telangiectasia, Ataxia with Oculomotor Apraxia, AOA, X-linked ataxia, FXTAS, HSP, hereditary spastic paraparesis, SPG, spastic paraplegia, spastic paraparesis, Strümpell-Lorrain, hereditary spastic spinal paralysis

2 Definitions per inheritance pattern

The suspected inheritance pattern was determined from the visits before a possible genetic cause was identified and was defined as followed: autosomal dominant (AD) if at least two generations were affected and there was evidence of parent to child transmission; autosomal recessive (AR) if only siblings showed similar symptoms or if the parents were consanguineous; X-linked (X) when only men were (more severely) affected and the trait was not passed from father to son; and mitochondrial when there was a maternal inheritance pattern and/or other characteristics as progressive external ophtalmoplegia, ptosis, diabetes, short stature, epilepsy. Patients were considered sporadic if no other family members from first, second or third degree were affected and no other inheritance pattern was suspected. If the family histories did not fit in one of these categories or if the family history was unsure, then an ‘ambiguous inheritance pattern’ was assigned.

3 Next generation sequencing: targeted exome approach

Sequencing libraries were generated after DNA shearing using the Kapa KAPA Hyperprep kit of Roche. The enriched fragments were then amplified and sequenced on a Illumina Hiseq2500, 2x 1 50bp paired-end sequencing. Samples were mapped using BWA on the reference genome hg19.[1] Duplicate reads were marked using Picard MarkDuplicates and base quality score recalibration was performed using GATK BaseRecalibrator. Variants were called using GATK[2] HaplotypeCaller. Variants were decomposed and left aligned using vt decompose and normalize commands and merged used bcftools merge command.[3] The merged vcf file was then annotated using VEP and loaded into a gemini database for querying.[4,5] For further filtering our custom algorithm selected variants with allele frequency less than 2% in existing databases including ExAC[6], Gnomad[7], 1000 Genomes[8], Exome Sequencing Project (ESP)[9],GoNL[10], Project MinE Belgian control samples [11] as well as in control samples in our in-house Cartagenia database. Furthermore, we only prioritized variants with non-low-impact score according to the IMPACT rating of the Ensembl browser[12] and/or variants previously annotated as ‘pathogenic’ or ‘likely pathogenic’ in Clinvar.[13]

Supplementary tables

ABCB7	ATXN10	DDHD1	GRM1	OPA1	SACS	TPP1
ABCD1	ATXN2	DDHD2	HACE1	PCNA	SAMD9L	TRPC3
ABHD12	ATXN3	DNAJC19	HSPD1	PDYN	SCARB2	TTBK2
ACO2	ATXN7	DNAJC3	IBA57	PEX10	SCYL1	TTPA
AFG3L2	ATXN8OS	DNMT1	IFRD1	PEX7	SETX	TUBB4A
ALDH18A1	B4GALNT1	DSTYK	ITPR1	PGAP1	SIL1	TWNK
ALDH3A2	BEAN1	EEF2	KCNA1	PHYH	SLC16A2	UCHL1
ALS2	BICD2	ELOVL4	KCNC3	PIK3R5	SLC1A3	USP8
AMPD2	BSCL2	ELOVL5	KCND3	PLD3	SLC33A1	VAMP1
ANO10	C12orf65	ENTPD1	KCNJ10	PLP1	SLC52A2	VLDLR
AP4B1	C19orf12	ERLIN1	KIF1A	PMM2	SLC9A1	VPS37A
AP4E1	CA8	ERLIN2	KIF1C	PMPCA	SNX14	VWA3B
AP4M1	CACNA1A	EXOSC3	KIF5A	PNKP	SPART	WASHC5
AP4S1	CACNA1G	FA2H	KLC2	PNPLA6	SPAST	WDR48
AP5Z1	CACNB4	FARS2	L1CAM	POLG	SPG11	WFS1
APTX	CAPN1	FAT2	LAMA1	POLR3A	SPG21	WWOX
ARL6IP1	CCDC88C	FGF14	MAG	PPP2R2B	SPG7	XRCC1
ARSI	CCT5	FLRT1	MARS	PRICKLE1	SPTBN2	ZFR
ATCAY	CLCN2	FLVCR1	MARS2	PRKCG	STUB1	ZFYVE26
ATG5	CLN5	FMR1	MME	PTF1A	SYNE1	ZFYVE27
ATL1	COQ8A	FXN	MRE11	PUM1	SYT14	ZNF592
ATM	CP	GBA2	MTPAP	RAB3GAP2	TBP	
ATN1	CPT1C	GBE	MTTP	REEP1	TDP1	
ATP13A2	CWF19L1	GDAP2	NIPA1	REEP2	TDP2	
ATP1A3	CYP27A1	GFAP	NKX6-2	RNF170	TECPR2	
ATP2B3	CYP2U1	GJC2	NOP56	RNF216	TGM6	
ATP8A2	CYP7B1	GOSR2	NPC1	RTN2	TK2	
ATXN1	DAB1	GRID2	NT5C2	RUBCN	TMEM240	

Supplementary table 1: ataxia-SPG panel
Gene panels have changed over time. Therefore, re-analysis of all data with the most recent gene panel (as depicted here) was performed by one of the authors, SVD in October,2021.

A Ataxia cohort

A1: Diagnoses in ataxia patients with age at onset of <27 years

Diagnosis (gene)	***Number of patients***	***Inheritance***	***% of total (n=75)***
Unexplained	36	-	48.0%
Friedreich ataxia *(FXN)*	21	AR	28.0%
Aoa2 *(SETX)*	3	AR	4.0%
Arsacs *(SACS)*	3	AR	4.0%
Spg7 *(SPG7)*	3	AR	4.0%
SCA7 *(ATXN7)*	3	AD	4.0%
SCAR16 *(STUB1)*	1	AR	1.3%
SCA3 *(ATXN3)*	1	AD	1.3%
EA2 *(CACNA1A)*	1	AD	1.3%
CTX *(CYP27A1)*	1	AR	1.3%
Ataxia and ophtalmoparesis *(POLG)*	1	AD	1.3%
Niemann-Pick disease *(NPC1)*	1	AR	1.3%

A2: Diagnoses in ataxia patients with age at onset of ≥27 years

Diagnosis (gene)	***Number of patients***	***Inheritance***	***% of total (n=174)***
Unexplained	114	-	65.5%
SCA6 *(CACNA1A)*	15	AD	8.6%
CANVAS *(RFC1)*	8	AR	4.6%
FXTAS *(FMR1)*	7	X	4.0%
SCA8 *(ATXN8)*	5	AD	2.9%
SCAR8 *(SYNE1)*	4	AR	2.3%
SCA7 *(ATXN7)*	3	AD	1.7%
SCA2 *(ATXN2)*	3	AD	1.7%
SCA3 *(ATXN3)*	2	AD	1.1%
SCA1 *(ATXN1)*	2	AD	1.1%
SCA17 *(TBP)*	2	AD	1.1%
EA2 *(CACNA1A)*	2	AD	1.1%
Friedreich ataxia *(FXN)*	2	AR	1.1%
AOA2 *(SETX)*	1	AR	0.6%
ADSA *(RNF170)*	1	AD	0.6%
SPG7 *(SPG7)*	1	AR	0.6%
SCAR10 *(ANO10)*	1	AR	0.6%
SPG10 *(KIF5A)*	1	AD	0.6%

B Spastic paraplegia cohort

B1: Diagnoses in SPG patients with age at onset of <25 years

Diagnosis (gene)	***Number of patients***	***Inheritance***	***% of total (n=59)***
Unexplained	33	-	55.9%
SPG4 *(SPAST)*	13	AD	22.0%
SPG15 *(ZFYVE26)*	3	AR	5.1%
SPG7 *(SPG7)*	2	AR	3.4%
SPG5 *(CYP7B1)*	1	AR	1.7%
SPG56 *(CYP2U1)*	1	AR	1.7%
ARSACS *(SACS)*	1	AR	1.7%
X chromosome structural abnormality[1]	1	X	1.7%
MASA syndrome *(L1CAM)*	1	X	1.7%
Silver syndrome *(BSCL2)*	1	AD	1.7%
HSP with *POLR3A* mutation	1	AR	1.7%
HSP with *TUBB4A* mutation	1	AD	1.6%

B2: Diagnoses in SPG patients with age at onset of ≥25 years

Diagnosis (gene)	***Number of patients***	***Inheritance***	***% of total (n=98)***
Unexplained	55	-	56.1%
SPG4 *(SPAST)*	29	AD	29.6%
SPG7 *(SPG7)*	7	AR	7.1%
SPG5 *(CYP7B1)*	2	AR	2.0%
SPG10 *(KIF5A)*	2	AD	2.0%
SPG31 *(REEP1)*	1	AD	1.0%
Polyglucosan body disease *(GBE1)*	1	AR	1.0%
Adrenomyeloneuropathy *(ABCD1)*	1	X	1.0%

Supplementary table 2: diagnoses in patients with a younger age at onset vs an older age at onset.
Based on the bimodal distribution of age at onset (*supplementary figure 1)*, the cohort was divided in a group of patients with younger age at onset (<27 years for ataxia patients (A1), <25 years for SPG patients (B1)) and a group of patients with older age at onset (≥27 years for ataxia patients (A2), ≥25 years for SPG patients (B2)). The frequency of the molecular diagnoses found in each group is demonstrated in the table.

[1] X-chromosome abnormality: karyotyping showed: 46,X,inv.dup(X)(q22.3q27.3). X-array-CGH (comparative genomic hybridization) showed a duplication of 47 Mb (Xq22.1-Xq28) and a deletion of 16 Mb (Xq28). Within the duplicated zone, a 0.8 Mb region was not duplicated. Further investigations showed skewed X-inactivation (>99%).

Abbreviations: AD, autosomal dominant; ADSA, autosomal dominant sensory ataxia; AOA2, ataxia-oculomotor apraxia type 2; AR, autosomal recessive; ARSACS, autosomal recessive spastic ataxia of Charlevoix-Saguenay; CANVAS, cerebellar ataxia, neuropathy, vestibular areflexia syndrome; CTX, cerebrotendinous xanthomatosis; EA2, episodic ataxia type 2; FXTAS, fragile X tremor/ataxia syndrome; HSP, hereditary spastic paraplegia; MASA, mental retardation, aphasia, shuffling gait and adducted thumbs; SCA, spinocerebellar ataxia; SCAR, spinocerebellar ataxia autosomal recessive; SPG, spastic paraplegia.

	Ataxia (n=262)		SPG (n=172)	
	# patients for whom test performed	# patients for whom test not performed	# patients for whom test performed	# patients for whom test not performed
SCA first series	218 Negative: 183 Positive: 35 *(16.1%)* - SCA1: 2 - SCA2: 4 - SCA3: 4 - SCA6: 19 - SCA7: 6	44[a]	17 Negative: 17 Positive: 0	155
SCA second series	106 Negative: 98 Positive: 7 *(6.6%)* - SCA8: 4 - SCA17: 3 Unsure: 1 - SCA17[b]	156[c]	2 Negative: 2 Positive: 0	170
FXN	155 Negative: 123 Positive: 28 *(18.1%)* Heterozygous: 4 *(2.6%)*	107	16 Negative: 16 Positive: 0	156
SPAST	8 Negative: 8 Positive: 0	254	154 Negative: 103 Positive: 51 *(33.1%)*	18[d]

Supplementary table 3: Number of patients for whom gene-by-gene testing was performed for SCA first series, SCA second series, FXN (Friedreich ataxia) and SPAST (SPG4).

[a]Of the 44 ataxia patients for whom the first SCA series had not been tested, 37 were known with another genetic explanation, mostly Friedreich ataxia (24) or FXTAS (7).
[b]One ataxia patient carried an intermediate repeat of unknown significance in *TBP* (SCA17), which was also present in the unaffected father (additional gene panel showed no pathogenic variants).
[c]Of the 156 patients for whom the second SCA series was not tested, 81 were known with another genetic diagnosis (mostly Friedreich ataxia (27)) or testing for the first SCA series had yielded a positive result (35).
[d]Of the 18 SPG patients for whom SPG4 was not tested, four were known with another diagnosis.

Abbreviations: SCA, spinocerebellar ataxia; SPG, spastic paraplegia.

A

	ATAXIA – SPG
Number of probands	128
Male (%)	68 *(53.1%)*
Ancestry	
European	113
African	0
Asian	7
Middle-East	3
America	2
Russia	0
Unknown	3
Age at onset *(years)*: mean [minimum-maximum; n]	31.7 [0-74; n=124]
Age at presentation *(years)*: mean [minimum-maximum; n]	43.2 [17-77; n=128]
Diagnostic delay (for patients explained by MPS) *(years)* : mean [minimum-maximum; n]	21.0 [4-56, n=36]
Inheritance pattern	
Suggestive inheritance pattern	63
AD history	19
AR history	10
X-linked history	0
Mitochondrial history	0
Ambiguous/uncertain history	34
Sporadic	65
Consanguinity	3
Phenotypic features	
Ataxia	68 *(53.1%)*
Spasticity	101 *(78.9%)*
Peripheral neuropathy	33 *(25.8%)*
Results	
Explained by MPS	37
Explained by other tests afterwards	3
Unexplained	88
Variants of interest found by MPS	29

B

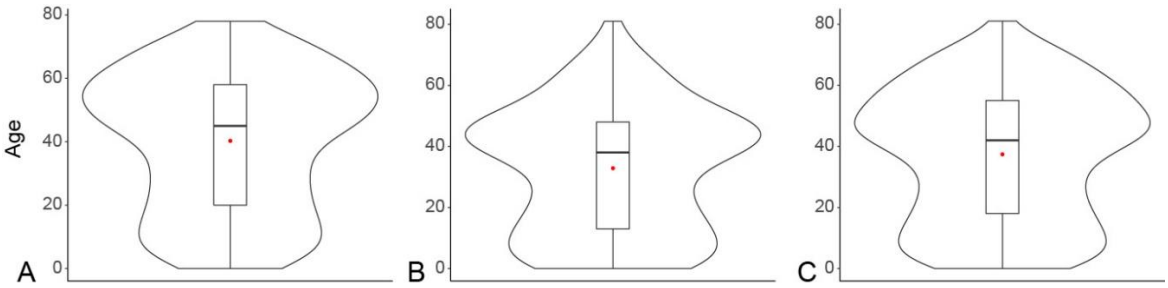
Number of patients for whom no gene-by-gene testing had been performed before performing a gene panel	3
Average number of gene-by-gene tests per patient (mean [min-max, n])	3.0 [0-11, n=128]
Number of patients for whom SCA first series was excluded	59
Number of patients for whom SCA second series was excluded	41
Number of patients for whom *FXN* was excluded	54
Number of patients for whom *SPAST* was excluded	70

Supplementary table 4: Characteristics of the cohort of ataxia and SPG patients for whom MPS was performed (n=128; 56 ataxia and 72 SPG).

A: Characteristics of the ataxia/SPG cohort for which MPS was performed
B: Gen-by-gene testing that was performed in patients that received a gene panel.

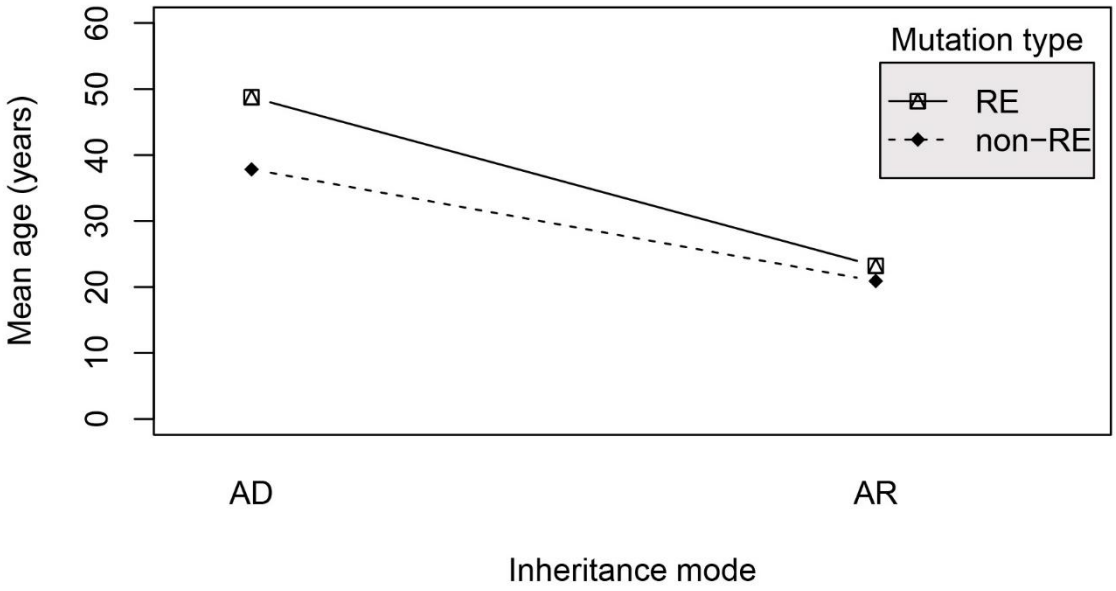
Abbreviations: AD, autosomal dominant; AR, autosomal recessive; MPS, massive parallel sequencing; n, total number of patients for whom data are available; SCA, spinocerebellar ataxia; SPG, spastic paraplegia.

Supplementary figures



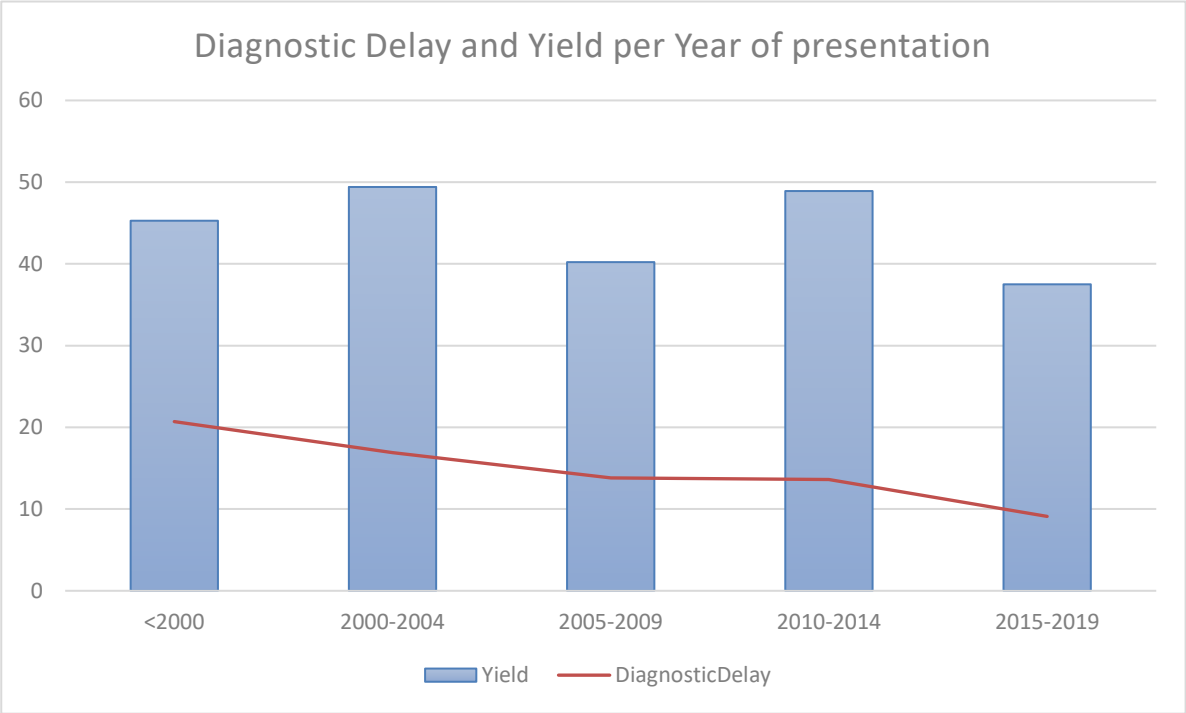
Supplementary figure 1: mean age at onset; violin plots.
Mean age at onset for **A**: ataxia only (n=249), **B**: SPG only (n=157), **C**: all patients with ataxia and SPG (n=406). Only patients for whom data are available are included in the analysis. A bimodal distribution of the age at onset is noted for ataxia as well as for SPG.The red dot denotes the mean. The horizontal line in the boxplots visualizes the median.

Abbreviations: SPG; spastic paraplegia



Supplementary figure 2: Interaction between inheritance mode and mutation type on age at onset.
Interaction plot to evaluate the interaction between the inheritance mode (AD or AR) and the mutation type (RE or non-RE) on the mean age at onset. Using Anova multivariate analysis, the interaction term was not significant (p=0.3).

Abbreviations: AD, autosomal dominant; AR autosomal recessive; RE, repeat expansion.



Supplementary figure 3: Evaluation of diagnostic delay and diagnostic yield depending of year of first presentation.

Chapter 5: Diagnostic yield of testing for *RFC1* repeat expansions in patients with unexplained adult-onset cerebellar ataxia

Sien H. Van Daele,[1,2,3] Sascha Vermeer,[4] Amélie Van Eesbeeck,[4] Laura Lannoo,[3] Valérie Race,[4] Philip Van Damme,[1,2,3] Kristl G. Claeys,[3,5] Wim Vandenberghe[3,6]

[1] KU Leuven – University of Leuven, Department of Neurosciences, Experimental Neurology, and Leuven Brain Institute (LBI), Leuven, Belgium

[2] VIB, Center for Brain & Disease Research, Laboratory of Neurobiology, Leuven, Belgium

[3] University Hospitals Leuven, Department of Neurology, Leuven, Belgium

[4] University Hospitals Leuven, Center for Human Genetics, Leuven, Belgium

[5] KU Leuven- University of Leuven, Laboratory for Muscle Diseases and Neuropathies, Department of Neurosciences, Experimental Neurology, Leuven, Belgium

[6] KU Leuven – University of Leuven, Department of Neurosciences, Experimental Neurology, Laboratory for Parkinson Research, Leuven, Belgium

Published in J Neurol Neurosurg Psychiatry 2020;91(11):1233-1234.

Introduction

Cerebellar ataxia, neuropathy, vestibular areflexia syndrome (CANVAS) is an adult-onset, slowly progressive neurodegenerative disorder characterized in its full form by cerebellar ataxia, sensory neuropathy and bilateral vestibulopathy. Autonomic dysfunction and dry cough are also frequent features. CANVAS is usually sporadic, but occasionally occurs in siblings. Recently, a biallelic intronic AAGGG repeat expansion $(AAGGG)_{exp}$ in *RFC1* was identified as the cause in almost all tested familial cases with full CANVAS syndrome.[1, 2] Interestingly, biallelic $(AAGGG)_{exp}$ in *RFC1* were also found in 22% of sporadic, previously idiopathic late-onset ataxia cases.[1] The percentage of positive cases increased to 63% if only sporadic patients with combined late-onset cerebellar ataxia and sensory neuropathy were considered, and to 92% in sporadic cases with full CANVAS.[1] In a cross-sectional description of 100 patients with biallelic *RFC1* $(AAGGG)_{exp}$, all cases had sensory neuropathy on nerve conduction studies (NCS), while only 79% of cases had cerebellar involvement.[3] Whether biallelic *RFC1* $(AAGGG)_{exp}$ can cause cerebellar ataxia without sensory neuropathy, remains uncertain. Here, we determined the diagnostic yield of testing for *RFC1* repeat expansions in a single-centre cohort of unexplained adult-onset ataxia cases.

Patients and methods

We included all index patients with ataxia who were evaluated at University Hospitals Leuven in the Movement Disorders Clinic between March 2005 and September 2019 or the Neuromuscular Clinic between January 2000 and December 2018 and fulfilled the following criteria: (1) onset age 18 years or older (onset as early as 19 years has been reported in an *RFC1* case) [3]; (2) ataxia as prominent clinical feature; (3) clinical evidence for cerebellar origin of the ataxia, that is. cerebellar oculomotor disorder (broken pursuit, dysmetric saccades) and/or cerebellar dysarthria; (4) no family history suggesting dominant inheritance; (5) no identified genetic or acquired cause and (6) no rapid progression typical of multiple system atrophy. We performed short-range flanking PCR and repeat-primed PCR for the $(AAAAG)_{exp}$, $(AAAGG)_{exp}$ and $(AAGGG)_{exp}$ *RFC1* alleles on genomic DNA using described primers.[1] The study was approved by the institutional Ethics Committee (approval number: S61184).

Results

We included 85 non-related index patients with unexplained adult-onset cerebellar ataxia (onset age: 52 ± 15.1 years (mean ± SD); 41 female). Nine of the 85 patients had an affected sibling, but these siblings were not included in the current analysis. All 85 probands had previously tested negative for spinocerebellar ataxia (SCA) type 1, 2, 3, 6 and 7. A negative genetic test had been obtained for SCA10, SCA12, SCA17 and dentatorubral-pallidoluysian atrophy in 62/85 patients, for Freidreich's ataxia in 61/85 and for Fragile X-associated tremor ataxia syndrome in 37/85.

We identified biallelic $(AAGGG)_{exp}$ in RFC1 in 9 patients (10.6%; five female; six sporadic). Clinical features are summarized in Table 1. Onset age was 56 ± 9.4 years. At the time of this study, five of nine patients had died (age at death 77 ± 5.9 years; mean survival from onset: 20 ± 5.5 years).

All nine positive cases had a neuropathy on NCS with predominant or exclusive sensory involvement. In eight of nine cases, this was associated with clinical symptoms and signs of sensory neuropathy. In three cases, NCS also showed motor involvement, although less prominent than the sensory component. Vestibular testing using video-oculography, caloric stimulation and a rotatory chair was performed in three of nine cases. All three had bilateral vestibular hypo- or areflexia and therefore full CANVAS syndrome (two sporadic, one familial). Six of nine cases had autonomic dysfunction and five of nine had chronic cough.

When we restricted our cohort to patients with both cerebellar ataxia and sensory neuropathy, the diagnostic yield of *RFC1* testing increased to 28.1% (9/32). If we only considered patients with ataxia onset after 35 years (similar to the cut-off for late-onset ataxia in a previous *RFC1* study[1]), the yield was 12.7% (9/71). If we only included patients with onset after 35 years and both cerebellar ataxia and sensory neuropathy, the yield was 30.0% (9/30).

Three of the nine positive cases had a sibling with reportedly similar symptoms. One of these (brother of patient 9) was available for evaluation and was found to also have full CANVAS and a biallelic $(AAGGG)_{exp}$ in *RFC1*. His clinical features were described in a recent case report [4] and he was not included in the determination of diagnostic yield, which was restricted to the 85 index patients.

Six of 85 patients (7.1%) were heterozygous for $(AAGGG)_{exp}$, which is considered non-pathogenic carrier status.[1] Three of them carried the reference $(AAAAG)_{11}$, two the $(AAAAG)_{exp}$ and one the $(AAAGG)_{exp}$ configuration on the other allele. The $(AAGGG)_{exp}$/$(AAAGG)_{exp}$ carrier had ataxia onset at 58 years and sensory neuropathy. Eight patients carried two repeat expansions of repeat units other than (AAGGG): biallelic $(AAAAG)_{exp}$ in four, biallelic $(AAAGG)_{exp}$ in one and $(AAAAG)_{exp}$/$(AAAGG)_{exp}$ in three.

Pt	Age at Onset (years), gender	Age at exa-mination (years)	Age at death (years)	Cerebellar atrophy on MRI	Downbeat nystagmus	Sensory symptoms /signs	Pyramidal symptoms/ signs	SNAP	CMAP	Vestibular Investi-gation	Autonomic	Chronic Cough	Familial
1	73, M	84	87	-	+	+	-	UL: ↓/absent LL: absent	UL: ↓ LL: ↓	NP	U	-	-
2	56, F	71	75	+	-	-	-	UL: ↓/absent LL: NP	UL: ↓ LL: ↓	NP	-	+	-
3	39, F	46	-	+	+	+	-	UL: ↓/absent LL: absent	UL: normal LL: normal	Bilateral vestibular hyporeflexia	U, OH	+	-
4	53, M	69	79	+	-	+	-	UL: ↓ LL: absent	UL: normal LL: normal	NP	-	+	-
5	49, M	59	74	+	-	+	B	UL: ↓ LL: absent	UL: normal LL: ↓	NP	U	-	Brother with similar symptoms (NA)
6	57, F	67	72	-	+	+	HR HT	UL: ↓ LL: absent	UL: normal LL: normal	NP	U	-	-
7	64, F	69	-	+	+	+	-	UL: ↓ LL: absent	UL: NP LL: normal	NP	-	+	Brother with similar symptoms (NA)
8	53, F	70	-	+	-	+	HR	UL: ↓ LL: absent	UL: normal LL: normal	Bilateral vestibular areflexia	Dry mouth	-	-
9	56, M	66	-	+	+	+	-	UL: absent LL: ↓/absent	UL: normal LL: normal	Bilateral vestibular areflexia	OH, I, C	+	Brother with similar symptoms

Table 1: Clinical characteristics of patient with biallelic pathogenic RFC1 expansions.

Abbreviations: B, Babinski sign; PT, patient ID; M, male; F, HR, hyperreflexia; HT, hypertonia; female; MRI, magnetic resonance imaging; UL, upper limbs; LL, lower limbs; NP, not performed; SNAP, sensory nerve action potential; CMAP, compound muscle action potential; U, urinary dysfunction; OH, orthostatic hypotension; I, impotence; C, constipation; NA, not available for evaluation.

Discussion

Biallelic $(AAGGG)_{exp}$ in *RFC1* were a relatively frequent cause of adult-onset cerebellar ataxia in our cohort. Importantly, all nine positive cases also had a sensory or sensorimotor neuropathy, although this was subclinical in one. None had pure cerebellar ataxia with normal findings on NCS. Based on these data, we recommend performing NCS in all cases of unexplained adult-onset cerebellar ataxia to detect a subclinical or confirm a clinically suspected sensory neuropathy. If NCS shows sensory neuropathy, *RFC1* testing is highly recommended. If NCS shows no sensory nerve impairment, *RFC1* mutations seem very unlikely. Consistent with this, a recent description of 100 patients with biallelic *RFC1* $(AAGGG)_{exp}$ identified sensory neuropathy as a common feature in all cases, while cerebellar ataxia was less frequent, leading the authors to suggest progression from sensory neuropathy early on to later cerebellar dysfunction.[3]

The frequency of positive cases in our adult-onset ataxia cohort was lower than in a London-based cohort, even if we used the same onset age cut-off of 35 years (12.7% in our study versus 22%),[1] but was higher than recently reported for adult-onset ataxia cohorts from Canada (4.3%) and Brazil (0.6%).[5] This may be due to differences in ethnic composition of the cohorts or different inclusion criteria for adult-onset ataxia.

We found one patient with cerebellar ataxia and subclinical sensory neuropathy and a $(AAGGG)_{exp}/(AAAGG)_{exp}$ state. This constitution was previously reported in a CANVAS patient,[2] but also in a healthy control.[1] Although we cannot exclude that the $(AAGGG)_{exp}/(AAAGG)_{exp}$ configuration might cause disease, evidence for its pathogenicity is currently insufficient. The $(AAGGG)_{exp}/(AAAAG)_{exp}$ configuration is also considered non-pathogenic.[1]

Limitations of our study were the lack of vestibular investigations in six of nine cases, the retrospective nature of clinical descriptions and the fact that we did not determine exact repeat expansion lengths. Longitudinal studies of biallelic *RFC1* $(AAGGG)_{exp}$ carriers from early to advanced stages will be required to fully characterize the clinical phenotype.

References

1. Cortese A, Simone R, Sullivan R, et al. Biallelic expansion of an intronic repeat in RFC1 is a common cause of late-onset ataxia. Nat Genet 2019;51(4):649-58.
2. Rafehi H, Szmulewicz DJ, Bennett MF, et al. Bioinformatics-Based Identification of Expanded Repeats: A Non-reference Intronic Pentamer Expansion in RFC1 Causes CANVAS. Am J Hum Genet 2019;105(1):151-65.
3. Cortese A, Tozza S, Yau WY, et al. Cerebellar ataxia, neuropathy, vestibular areflexia syndrome due to RFC1 repeat expansion. Brain 2020;143(2):480-90.
4. Terryn J, Van Eesbeeck A, Vermeer S, et al. The characteristic eye movement disorder of RFC1-linked CANVAS. Mov Disord Clin Pract 2020;7(2):230-1.
5. Akcimen F, Ross JP, Bourassa CV, et al. Investigation of the RFC1 Repeat Expansion in a Canadian and a Brazilian Ataxia Cohort: Identification of Novel Conformations. Front Genet 2019;10:1219.

Chapter 6: *RNF170* mutation causes autosomal dominant sensory ataxia with variable pyramidal involvement

Sien H. Van Daele MD [1,2,3], Matthieu Moisse PhD [1,2], Valérie Race PhD[4], Amélie Van Eesbeeck[4], Liesbeth Keldermans[4], Sascha Vermeer MD PhD[4], Hilde Van Esch MD PhD[4,5], Kristl G. Claeys MD PhD [3,6], Philip Van Damme MD PhD [1,2,3]

[1] KU Leuven – University of Leuven, Department of Neurosciences, Experimental Neurology, and Leuven Brain Institute (LBI), Leuven, Belgium

[2] VIB, Center for Brain & Disease Research, Laboratory of Neurobiology, Leuven, Belgium

[3] University Hospitals Leuven, Department of Neurology, Leuven, Belgium

[4] University Hospitals Leuven, Center for Human Genetics, Leuven, Belgium

[5] KU Leuven- University of Leuven, Laboratory for the Genetics of Cognition, Department of Human Genetics, Leuven, Belgium.

[6] KU Leuven- University of Leuven, Laboratory for Muscle Diseases and Neuropathies, Department of Neurosciences, and Leuven Brain Institute (LBI), Leuven, Belgium.

Published in Eur J Neurol 2022; 29(1):345-349.

Abstract

BACKGROUND: Although hereditary ataxias are a group of clinically and genetically heterogeneous disorders, specific clinical clues can sometimes incriminate certain genes. This can trigger genetic testing in sporadic patients or prompt dissecting certain genes more thoroughly when initial genetic testing is negative. Also for the assembly of gene panels and interpretation of the results, genotype-phenotype correlations remain important to establish.

METHODS: We clinically evaluated a Belgian family with autosomal dominant inherited sensory ataxia and variable pyramidal involvement and performed targeted clinical exome sequencing. Secondly, we retrospectively screened sequencing data of an in-house cohort of 404 patients with neuromuscular disorders for variants in the identified gene *RNF170*.

RESULTS: All affected family members showed sensory ataxia on examination. Pyramidal involvement, and sometimes slow-pursuit abnormalities and/or a sensory neuropathy were more variable findings. We identified the heterozygous variant p.Arg199Cys in *RNF170* in all three affected siblings of our family. We did not find additional pathogenic variants in *RNF170* in our in-house neuromuscular cohort.

CONCLUSION: We confirm the heterozygous variant p.Arg199Cys in *RNF170* in a Belgian family with autosomal dominant sensory ataxia and variable pyramidal involvement. This constitutes a rare but clinically recognizable phenotype that warrants testing of *RNF170*. Unlike the distinctive bi-allelic loss of function variants in *RNF170* associated with HSP, the p.Arg199Cys variant is the only one reported in sensory ataxia. It is important for neurologists to be aware of this characteristic phenotype and to include this gene in gene panels for ataxia and HSP.

Introduction

Hereditary ataxias (HA) are a group of clinically and genetically heterogeneous disorders, for which >100 genes have been associated.[1] Clinical presentation is often non-specific making it nearly impossible to deduce the underlying molecular deficit from the patient's phenotype. Therefore, next generation sequencing is an invaluable diagnostic tool. In some cases, however, clinical clues can point in the direction of certain genes, as is the case for *SPG7* and *SACS*. This can trigger genetic testing in sporadic patients or prompt dissecting certain genes more thoroughly when initial genetic testing is negative. Also for the assembly of gene panels and their analysis, awareness of genotype-phenotype correlations remains of great importance. We describe the clinically recognizable phenotype of patients with a specific heterozygous *ring finger protein 170* (*RNF170)* mutation and evaluate the presence of variants in this gene in a local cohort of 404 patients with neuromuscular disorders.

Methods

Clinical Case

We clinically evaluated the affected members of a family with an autosomal dominant (AD) history of sensory ataxia. We performed targeted clinical exome sequencing on the proband using DNA extracted from blood lymphocytes after written informed consent. We analyzed a gene panel containing 187 genes known to be associated with HA and/or hereditary spastic paraplegia (HSP). Variants were classified according to the American college of medical genetics and genomics (ACMG)-criteria.[2] We used Sanger sequencing to confirm the identified *RNF170* variant in the proband and for segregation analysis in the affected siblings.

Retrospective analysis of in-house cohort

To establish the frequency of (other) pathogenic *RNF170* variants, we retrospectively screened *RNF170* in an in-house cohort of 404 non-related patients with neuromuscular disorders (onset age 31 +- 18 years [mean ± standard deviation]; 167 female) for whom the same gene panel had been performed. This study was approved by the Medical Ethics Committee UZ KU Leuven (S64335).

Results

Case description

A 66-year-old Belgian woman (III.2) had experienced a sense of fatigue in the legs since the age of 35 years, evolving to an uncoordinated gait pattern and reduced sensation in the feet at the age of 60 years (Figure 1). Neurological examination showed manifest sensory loss, ataxia, diminished reflexes and hypertonia with Babinski signs in the lower limbs and discrete ataxia in the upper limbs. Gait was spastic-ataxic, Romberg sign was present and smooth eye pursuit mildly jerky. No nystagmus nor dysarthria was noted. Somatosensory evoked potentials (SSEPs) were absent in the lower extremities. Nerve conduction studies (NCS) and Electromyography (EMG) revealed no signs of neuropathy. Magnetic resonance imaging (MRI) full spine was normal (Supplementary figure 1), brain MRI showed limited corticosubcortical frontal atrophy. Metabolic work-up was normal. Friedreich's ataxia, spinocerebellar ataxia 1,2,3,6,7,8,10,12,17 and dentatorubral-pallidoluysian atrophy were excluded by targeted gene testing.

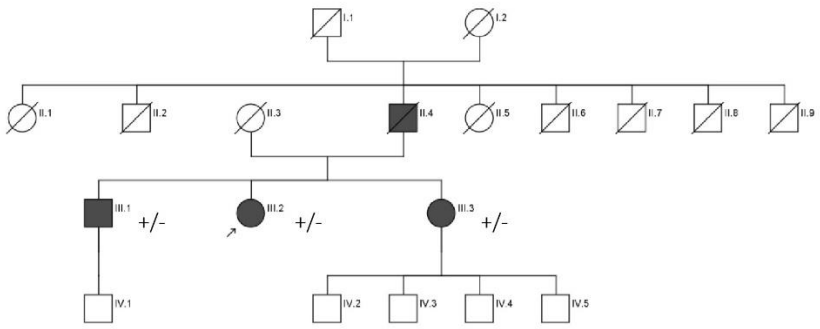
The patient's father (II.4), her older brother (III.1) and younger sister (III.3) had experienced similar symptoms of progressive gait instability and sensory loss in the lower limbs since the age of 64, 68 and 55 respectively (Figure 1). The siblings displayed ataxia, reduced reflexes and sensory disturbances in the legs on examination. Past examination of the father, who died at the age of 84, had shown similar findings along with Babinski signs and progressive hypertonia in the lower limbs and mild ataxia in the upper limbs. SSEPs were abnormal for both siblings. NCS/EMGs were normal for the father and sister but showed reduced sensory nerve action potentials (SNAPs) for the brother, compatible with axonal sensory neuropathy. The MRI of the spine of the father was normal, but it was not performed in the siblings.

Analysis of the gene panel for the proband showed a likely pathogenic missense variant in *RNF170* (ENST00000534961.1:c.595C>T; p.Arg199Cys; g.42711484G>A), confirmed by Sanger sequencing. No second variant in *RNF170* was detected (Supplementary table 1). The proband's affected siblings both harbored the same variant. Blood samples of the parents were not available.

Retrospective analysis of in-house cohort

In our local cohort of 404 patients with neuromuscular disease, we identified seven variants in 13 patients (Supplementary table 2), none was considered disease-causing. No patient carried two variants in *RNF170* (Supplementary table 1).

A



B

	Index pt (III.2)	Father (II.4)	Brother (III.1)	Sister (III.3)	Canadian families (Valdmanis et al) [3,4]	Ecuadorian family (Cortese et al) [6]
Age of onset *(years)*	35	64	68	55	28-80	47-60
Ataxia	LL>>>UL	LL>>>UL	LL	LL>>UL	LL>>>UL	LL>>>UL
Reflexes	Reduced in UL Absent in LL	Normal in UL Absent LL	Absent in UL Absent in LL	Normal in UL Reduced in LL	Reduced, absent or normal in UL Reduced to absent in LL	Reduced/normal in UL Absent in LL
Pyramidal signs	Increased muscle tone and bilateral Babinski signs in LL	Increased muscle tone and bilateral Babinski signs in LL	-	-	Some patients with extensor plantar response	-
Sensory loss	LL	LL	LL	LL	LL>>>UL	LL>>>UL
Eye movements	Jerky	Normal	Normal	Normal	Jerky in some	Normal
SSEP	Absent LL	Not done	Absent LL	Absent left leg Abnormal right leg	Absent LL	Not done
NCS	Normal	Normal	Reduced SNAPs in UL and LL	Normal	SNAPs sometimes impaired in older patients	Reduced/absent SNAPs in UL Absent SNAPs in LL
MRI brain	Corticosubcortical atrophy	Not done	Not done	Not done	Normal	Normal
MRI spine	Normal	Normal	Not done	Not done	Normal	Reduced volume and increased T2 signal of the posterior columns
Additional findings	-	-	-	-	-	Bilateral vestibular areflexia in the proband

Figure 1: Pedigree and clinical characteristics of patients with the RNF170 p.Arg199Cys mutation.
A: Pedigree of the family with sensory ataxia described in this paper. The proband is indicated with an arrow. A '+/-'-sign indicates patients that are tested and found heterozygous for the p.Arg199Cys variant. **B**: Clinical findings in this family as well as core features of the two Canadian families and the Ecuadorian family with the RNF170 p.Arg199Cys mutation.[3-6]

Abbreviations: LL, lower limbs; MRI, magnetic resonance imaging; NCS, nerve conduction studies; SNAP, sensory nerve action potential; SSEP, somatosensory evoked potential; UL, upper limbs.

Discussion

In a European family with AD late-onset sensory ataxia, we identified the heterozygous p.Arg199Cys variant in *RNF170*. This variant has been described before in two Canadian families and an Ecuadorian family, in which the affected members showed a very similar phenotype of sensory ataxia, variable pyramidal involvement and sometimes slow-pursuit abnormalities and/or sensory neuropathy (Figure 1).[3-6] Interestingly, the heterozygous p.Arg199Cys mutation is the only variant reported in *RNF170* for patients with this phenotype and no other heterozygous variants have yet been classified as pathogenic in this gene. In our in-house cohort, we did find a patient who carried a different heterozygous variant at the same position (ENST00000534961.1:c.596G>A; p.Arg199His; g.42711483C>T) but the frequency in population databases was too high to fit with the AD inheritance pattern in the family and the phenotype was different from the one of our proband.[7]

This indicated that p.Arg199His is not pathogenic. A possible explanation for the different effect of these two identically located amino acid substitutions is the different physiochemical properties of the substituting residues. Studies in mice have suggested that the positive charge of the residue at position 199 is crucial for the stability of the protein because of ionic interactions with amino acids located in transmembrane domains.[8] The other six *RNF170* variants in our cohort were also not considered pathogenic.

The reason why only the heterozygous p.Arg199Cys variant has been linked to the phenotype and this in several apparently unrelated families remains unclear. The two Canadian families shared the disease haplotype, indicating a founder mutation.[4] However, a common ancestry seems less likely for the other, geographically scattered families although a shared genetic background cannot be excluded since haplotype analysis was not performed. Given the lack of genetic variation in the immediate vicinity in population databases, except for the presumably benign p.Arg199His variant discussed earlier, the region does not seem to be a mutational hotspot.[7] Interestingly, amino acid residues important for the ionic interactions between transmembrane domains in *RNF170* mentioned above (Arg199, Arg201, Asp232, Asp233) are well conserved across species and their positions show lack of non-synonymous variants in population databases (Supplementary figure 2*).*[7,8] We hypothesize that (charge-changing) mutations of these residues may produce a similar phenotype as the one associated with the p.Arg199Cys mutation. In our family, the absence of symptoms in the paternal grandparents (I.1 and I.2) (both died at age >70 years), aunts and uncles (as far as known by our proband), suggests the possibility of a *de novo* mutation in the father although reduced penetrance cannot be excluded.

Bi-allelic mutations in *RNF170* cause early-onset HSP.[9,10] In contrast to the unique heterozygous p.Arg199Cys mutation in patients with sensory ataxia, distinct bi-allelic mutations have been described in HSP families. Although there is some minor overlap between the two phenotypes (sometimes subclinical impaired SSEPs in HSP patients, pyramidal involvement in sensory ataxia patients and (mild) cerebellar signs in both), it appears that *RNF170* mutations cause two distinct phenotypes with different inheritance through different mechanisms. For the heterozygous p.Arg199Cys variant, located in the transmembrane region of the protein, a gain of function mechanism is suspected.[5,8] For the bi-allelic variants in HSP, the mechanism is presumably loss of function.[9] Heterozygous carriers in these families have been found unaffected.[9] The different reported mutations in *RNF170* and their relation to phenotype, gene structure and protein domains are shown in Figure 2.

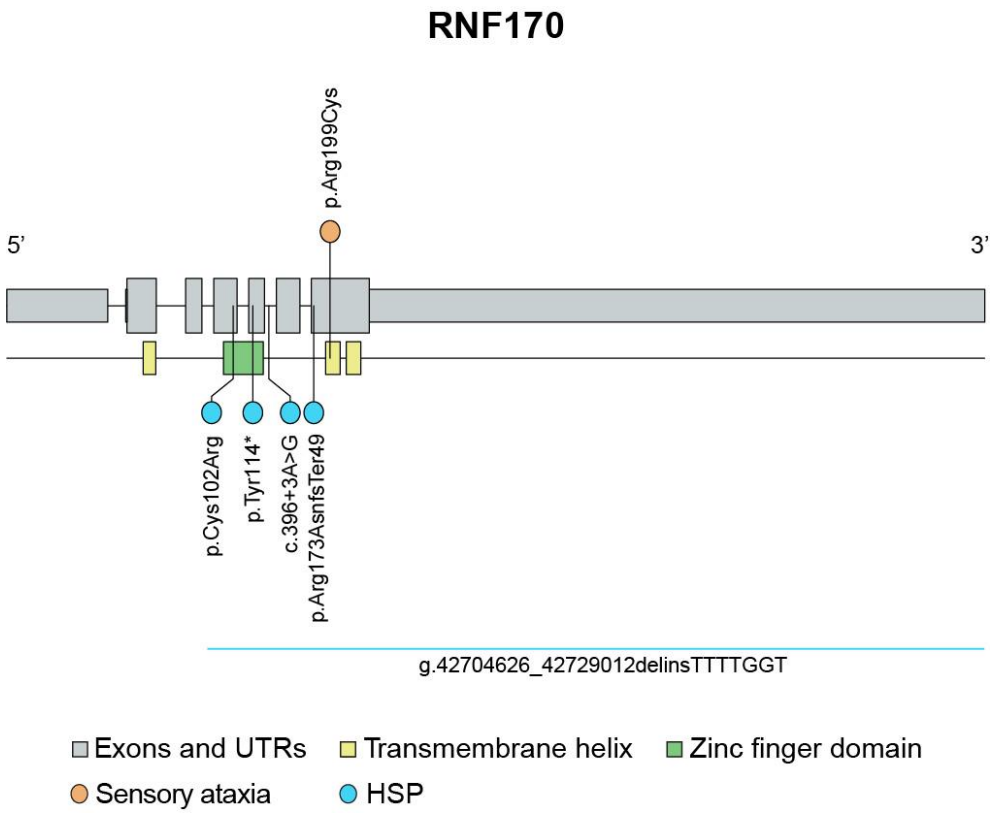


Figure 2: Mutational landscape of phenotypes related to RNF170.
The figure shows a lolliplot of gene *RNF170* locating the heterozygous variant described in sensory ataxia (*orange*) and the homozygous variants in HSP (*blue*) (including the large deletion in HSP shown as a blue line) in relation to gene structure and protein domains.[9,10] Introns are scaled with a factor 1/100.

Abbreviations: HSP, hereditary spastic paraplegia; UTR, untranslated region.

Given the characteristic clinical picture of sensory ataxia and variable pyramidal involvement in patients with a heterozygous *RNF170* mutation, clinicians can recognize this phenotype and ensure testing for *RNF170*. Especially for patients with ataxia and AD inheritance pattern, careful clinical characterization is important since sensory ataxia is often mistaken for cerebellar ataxia.[11] This confusion would cause clinicians to prioritize screening for repeat expansion disorders and even lead to failure of obtaining the correct genetic diagnosis in the end. Gene panels for ataxia can therefore best include genes associated with sensory ataxia such as *RNF170* and should also be performed in patients with AD family history, after exclusion of repeat expansions when ataxia is mainly cerebellar. Also in gene panels for HSP, *RNF170* should evidently be included.

Other possible genetic causes of sensory ataxia include Friedreich's ataxia, ataxia with vitamin E deficiency, abetalipoproteinemia, spinocerebellar ataxia with neuropathy, Charcot-Marie-Tooth disease 4C, "Cerebellar ataxia, neuropathy, vestibular areflexia syndrome" (CANVAS) and mitochondrial neuropathies, but often other features (and/or cerebellar ataxia) dominate the clinical picture in these patients.[6,11,12]

In summary, although patients with HA often show non-specific and highly variable phenotypes due to genetic and phenotypic heterogeneity, patients with a heterozygous *RNF170* mutation exhibit a clinically recognizable phenotype of sensory ataxia, variable pyramidal involvement and sometimes slow-pursuit abnormalities and/or sensory neuropathy. SSEPs are typically impaired and SNAPs often well preserved, compatible with preganglionic pathology. We confirm the relation between the specific heterozygous p.Arg199Cys mutation and sensory ataxia for the first time in a Belgian family. Neurologists should be vigilant for this phenotype and thus enable testing for *RNF170*, nowadays most commonly as part of a gene panel for ataxia/HSP.

References

1. Jayadev S, Bird TD. Hereditary ataxias: overview. *Genet Med.* 2013;15(9):673-683. doi:10.1038/gim.2013.28.
2. Richards S, Aziz N, Bale 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.* 2015;17(5):405-424. doi:10.1038/gim.2015.30.
3. Valdmanis PN, Simoes Lopes AA, Gros-Louis F, Stewart JD, Rouleau GA, Dupre N. A novel neurodegenerative disease characterised by posterior column ataxia and pyramidal tract involvement maps to chromosome 8p12-8q12.1. *J Med Genet.* 2004;41(8):634-639. doi:10.1136/jmg.2004.019711.
4. Valdmanis PN, Brunet D, St-Onge J, Weston L, Rouleau GA, Dupre N. A founder haplotype for autosomal dominant sensory ataxia in Eastern Canada. *Neurology.* 2006;67(12):2239-2242. doi:10.1212/01.wnl.0000249314.96183.48.
5. Valdmanis PN, Dupre N, Lachance M, et al. A mutation in the RNF170 gene causes autosomal dominant sensory ataxia. *Brain.* 2011;134(Pt 2):602-607. doi:10.1093/brain/awq329.
6. Cortese A, Callegari I, Curro R, et al. Mutation in RNF170 causes sensory ataxic neuropathy with vestibular areflexia: a CANVAS mimic. *J Neurol Neurosurg Psychiatry.* 2020;91(11):1237-1238. doi:10.1136/jnnp-2020-323719.
7. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature.* 2020;581(7809):434-443. doi:10.1038/s41586-020-2308-7.
8. Wright FA, Lu JP, Sliter DA, Dupre N, Rouleau GA, Wojcikiewicz RJ. A Point Mutation in the Ubiquitin Ligase RNF170 That Causes Autosomal Dominant Sensory Ataxia Destabilizes the Protein and Impairs Inositol 1,4,5-Trisphosphate Receptor-mediated Ca2+ Signaling. *J Biol Chem.* 2015;290(22):13948-13957. doi:10.1074/jbc.M115.655043.
9. Wagner M, Osborn DPS, Gehweiler I, et al. Bi-allelic variants in RNF170 are associated with hereditary spastic paraplegia. *Nat Commun.* 2019;10(1):4790. doi:10.1038/s41467-019-12620-9.
10. de Sainte Agathe JM, Mercier S, Mahe JY, et al. RNF170-Related Hereditary Spastic Paraplegia: Confirmation by a Novel Mutation. *Mov Disord.* 2021;36(3):771-774. doi:10.1002/mds.28371.
11. Chhetri SK, Gow D, Shaunak S, Varma A. Clinical assessment of the sensory ataxias; diagnostic algorithm with illustrative cases. *Pract Neurol.* 2014;14(4):242-251. doi:10.1136/practneurol-2013-000764.
12. Gooding R, Colomer J, King R, et al. A novel Gypsy founder mutation, p.Arg1109X in the CMT4C gene, causes variable peripheral neuropathy phenotypes. *J Med Genet.* 2005;42(12):e69. doi:10.1136/jmg.2005.034132.

Supplementary data

Supplementary tables

ID	% GQ > 70	missing pieces *RFN170*
Proband	95,74	54bp in exon 3; 42bp in exon 6
Patient1	95,18	3bp in exon 7
Patient2	93,05	0bp
Patient3	94,92	0bp
Patient4	NA	NA
Patient5	94,59	0bp
Patient6	96,55	0bp
Patient7	96,76	0bp
Patient8	97,04	0bp
Patient9	96,72	0bp
Patient10	98,96	0bp
Patient11	97,70	3bp in exon 7
Patient12	99,61	0bp
Patient13	99,00	0bp

Supplementary table 1: %GQ>70 (fraction of positions in the clinical target with a genotype quality above 70) and missing pieces in RNF170 sequencing data of the targeted clinical exome performed.
Data are shown for the proband of our family and for the 13 patients for whom a variant in *RNF170* was found in retrospective analysis of the sequencing data. The fraction of positions in the clinical target with genotype quality above 70 was high and there were few missing pieces, meaning that the gene was well covered and we do not expect to have missed second variants in *RNF170*.

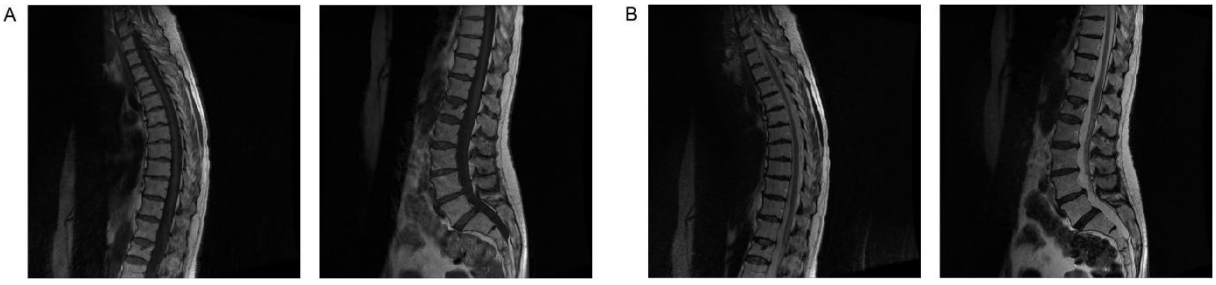
Abbreviations: bp, base pairs; NA, not available; GQ, genotype quality.

Variant (GRCh37)	cDNA	Protein	Zyg	ACMG classi-fication	ACMG-Criteria used	Transcript	HA *(n=49)*	HSP *(n=56)*	OTHER *(n=299)*
g.42711553insC	c.526dup	p.Asp176Glyfs*47	HTZ	VUS	PM2, PP3	ENST00000534961.1	/	/	1
g.42720581C>T	c.374G>A	p.Ser125Asn	HTZ	LB	BS1, BP4, BP6	ENST00000534961.1	/	/	6
g.42706017del	c.520del	p.His174Metfs*17	HTZ	VUS	PM2	ENST00000319104.3	/	/	2
g.42711333C>T	c.746G>A	p.Arg249Gln	HTZ	VUS	BS1, PP3	ENST00000534961.1	/	/	1
g.42705986C>T	c.551G>A	p.Arg184His	HTZ	LB	BS1, BP4	ENST00000319104.3	/	/	1
g.42742920C>T	c.88G>A	p.Val30Ile	HTZ	VUS	BS1, PP3	ENST00000534961.1	/	1	/
g.42711483C>T	c.596G>A	p.Arg199His	HTZ	VUS	PM5, PP3	ENST00000534961.1	/	/	1

Supplementary table 2: Variants in RNF170 in a cohort of 404 patients with neuromuscular diseases.
Seven variants in *RNF170* were found in 13/404 patients with neuromuscular diseases from our in-house cohort. Genomic position, cDNA-position, protein position and zygosity of the variant are displayed, along with the respective transcript. The result of the applied ACMG-criteria is shown as well as the criteria used. The last three columns show in how many patients with a particular phenotype (HA, HSP or other neuromuscular disorders) the variant has been identified.

Abbreviations: ACMG, American college of medical genetics and Genomics; cDNA, complementary DNA; GRCh37, genome reference consortium human build 37; HA, hereditary ataxia; HSP, hereditary spastic paraparesis; LB, likely benign; n, number; HTZ, heterozygous; VUS, variant of unkown significance; Zyg, zygosity.

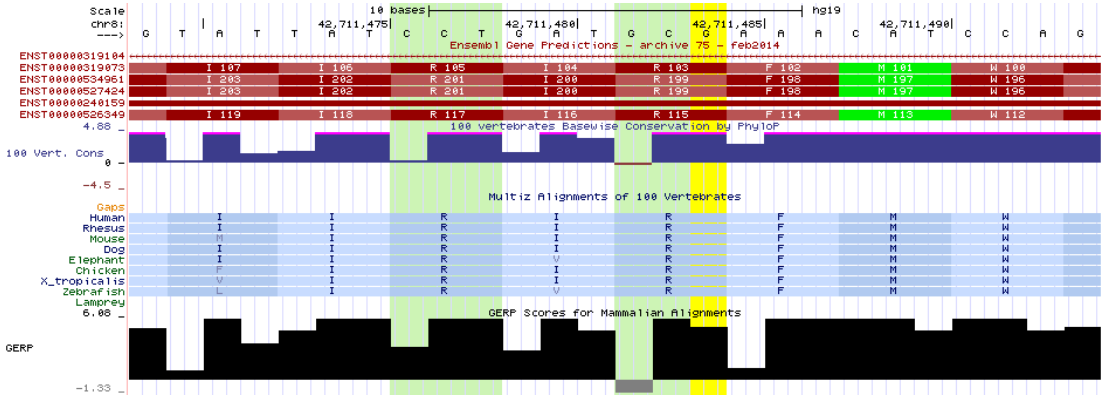
Supplementary figures



Supplementary figure 1: MRI spine of the proband (only sagittal images are available).
(A) T1 sagittal and (B) T2 sagittal images of the cervicothoracic and thoracolumbar spine show no abnormalities.

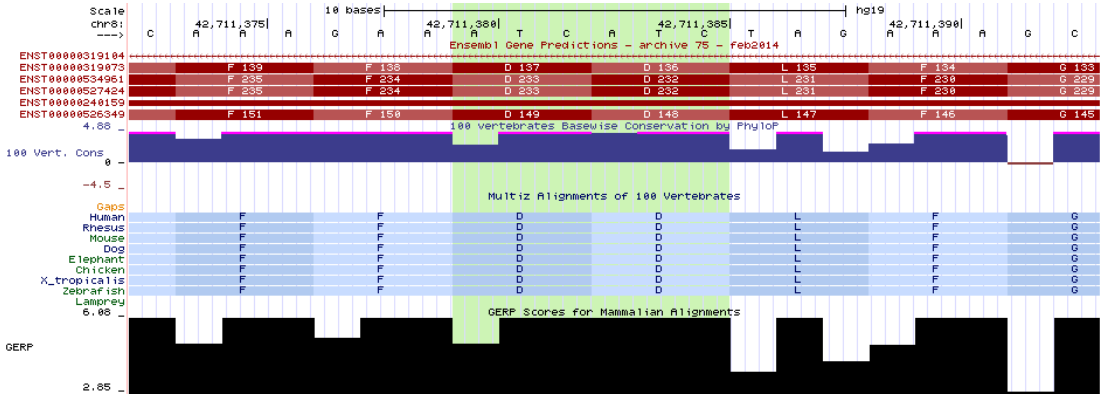
A

A1



From: https://genome.ucsc.edu/. GRCh37/hg19 assembly.

A2



From: https://genome.ucsc.edu/. GRCh37/hg19 assembly.

B

Genomic position	Protein position	GERP++_rs-score
chr8:42711482	Arg199	-1.33
chr8:42711483	Arg199	6.08
chr8:42711484	Arg199	5.2
chr8:42711476	Arg201	3.24
chr8:42711477	Arg201	5.99
chr8:42711478	Arg201	6.08
chr8:42711383	Asp232	6.08
chr8:42711384	Asp232	6.08
chr8:42711385	Asp232	6.08
chr8:42711380	Asp233	4.9
chr8:42711381	Asp233	6.08
chr8:42711382	Asp233	6.08

Supplementary figure 2: Conservation of the residues critical for ionic interactions in RNF170: (A) Conservation of amino acids and (B) GERP++ scores [2].

A1: Positions of interest (ENST00000534961.1; Arg199, Arg201) are highlighted in green. Variant ENST00000534961.1 c.595C>T; p.Arg199Cys; g.42711484G>A is highlighted in yellow. Conservation scores PhyloP [1] (100 Vert. cons) and GERP [2] are displayed. Amino acid sequences of different vertebrates are shown.

Note: the fact that the third nucleotide of the codon for Arg199 and Arg201 has a lower PhyloP and GERP score can be explained by the fact that all possible combinations of nucleotides (A, C, G, T) on the third position of the codon code for Arginine, when the first two remain the same (CG).

A2: Positions of interest (ENST00000534961.1; Asp232, Asp233) are highlighted in green.Conservation scores PhyloP [1] (100 Vert. cons) and GERP [2] are displayed. Amino acid sequences of different vertebrates are shown.

Chapter 7: General discussion and perspectives

In this dissertation, we aimed to gain more insight in the genetic determinants of disorders affecting the motor system. Therefore, we studied three disorders with an important genetic basis that mainly affect the motor system: hereditary ataxia, HSP and ALS.

MPS and gene panel analysis in disorders affecting the motor system

Yield and genotype-phenotype correlations

With gene panel analysis for ataxia and SPG, we obtained a yield of 28.9%, which is comparable to the diagnostic rates of other studies performing MPS for these disorders.[53, 54, 65, 94-107] However, as explained in the introduction, a true comparison between studies is difficult because of heterogeneity in the study designs. For example, in many of these studies only specific subpopulations were investigated (e.g. only adult or pediatric onset, only familial or sporadic patients and often only SPG or ataxia patients); a different number of genes were excluded by gene-by-gene testing prior to inclusion and the composition of the gene panels was different among distinct research centers. Since we did not use inclusion criteria based on family history or age of onset, we provided an estimation of the diagnostic rate that can be expected in an unselected cohort in real-world clinical practice. It is true that in routine clinical practice not only patients with a high suspicion of a genetic etiology are offered MPS but also otherwise unexplained ataxia/SPG patients without family history and/or late onset are often screened for underlying genetic causes. Although the expected yield is lower, it is known that some of these cases harbor pathogenic variants explaining their phenotype, as has been described in our and other studies.[41, 109] This validates the approach to search for a genetic etiology in similar patients presenting at the clinic, even if a priori probabilities are not as high as for patients with younger onset and clear familial history. Furthermore, it seems reasonable to pool ataxia and SPG together to estimate the yield of gene panel analysis. Although the typical phenotypes at the end of the spectrum undoubtedly have their own genetic profile, both phenotypic and genetic overlap (as described in the introduction) make it seem artificial to split this group in two distinct phenotypes.

With our gene panel analysis, we identified most pathogenic variants in *SPG7*, followed by *SETX*, *SYNE1* and *SACS* for ataxia patients and in *SPG7*, followed by *CYP7B1*, *REEP1*, *ZFYVE26* and *KIF5A* for SPG patients. Especially variants in *SPG7* are a common finding in other studies in mainly Caucasian populations.[54, 65, 75, 94, 96, 102, 103] Mutations in *SETX*, *SYNE1* and *SACS* are also a relatively frequent finding in ataxia patients and mutations in *CYP7B1*, *REEP1* and *ZFYVE26* in SPG patients, although results for these genes are more variable.[54, 75, 94, 102, 103] The remaining identified genes in our and other studies are very diverse and underscore the genetic heterogeneity in these disorders. It underlines the importance of the application of MPS, making it

possible to evaluate a large set of genes simultaneously. Given the clinical heterogeneity, it is a great advantage that MPS encompasses a hypothesis-free approach that is not restricted to previously known genotype-phenotype correlations. A remark here is that this hypothesis-free aspect is partly lost when a gene panel approach is used. Though, it still remains important. First, there is the possibility to broaden the scope – at least if (targeted) WES is used – and to consequently look at genes not yet associated with the phenotype when the case remains unexplained after initial gene panel analysis. Secondly, even when the analysis is restricted to the gene panel associated with the phenotype, multiple genes are considered, independent of the exact phenotype (since even between different patients with e.g. ataxia, the phenotype can differ e.g. depending on the associated symptoms beside ataxia). This means that there is still 'less hypothesis involved' compared to gene-by-gene testing. Some genes would never be considered for patients with a certain phenotype before the advent of MPS.

However, on the other hand, establishing genotype-phenotype correlations remains particularly important. This may seem somewhat paradoxical at first sight, but can be appreciated if we look more closely at the procedure of analyzing MPS data in a diagnostic setting. Since multiple variants are typically encountered per patient, multiple lines of evidence must be used to assess if one of these variants is likely to be pathogenic and responsible for the phenotype (as is illustrated by the ACMG-criteria for the interpretation of variants).[92] Known genotype-phenotype correlations are a very important aspect during variant analysis. Even though a variant might be suspicious on a purely genetic level, it is not advisable to report a variant in a certain gene as pathogenic when the phenotype of the patient does not match with the clinical characteristics described in literature for this gene. As a result, such variants are often classified as 'variants of unknown significance' (VUS). Therefore it remains important to further establish genotype-phenotype correlations in the era of MPS, both for correct variant interpretation in diagnostic setting as for a comprehensive assembly of gene panels. In the work performed for this thesis we contributed to this field, especially in chapter 4, 5 and 6. In chapter 4 we described genotype-phenotype correlations for the pathogenic variants and variants of interest that we encountered in our cohort of ataxia and SPG patients. It can serve as a reference work for genetic laboratories encountering the same (or similar) variants in their patients. In chapter 5, we examined genotype-phenotype correlations for the recently discovered *RFC1* biallelic AAGGG-expansions in patients with CANVAS. Before the discovery of the underlying molecular defect, this syndrome was thought to almost invariably present with the typical triad of cerebellar ataxia, (mainly) sensory neuropathy and bilateral vestibular hypofunction. However, following the identification of *RFC1* repeat expansions, we and others found that (subclinical) sensory neuropathy seems to be the core clinical feature in these patients, while ataxia and vestibular involvement are more variable characteristics. These findings learn us that all patients with sensory neuropathy (either as primary phenotypic feature or subclinical in combination with

cerebellar ataxia) should be screened for *RFC1* repeat expansions. More recent work indicates that patients with *RFC1* repeat expansions can also present with a multiple system atrophy (MSA)-like phenotype and that rarely, these patients have normal sensory nerve action potentials (SNAPs) on nerve conduction studies. However, in these patients only smaller repeat lengths (<400 repeats) are found and it remains to be proven that these are truly pathogenic.[110, 111] All this illustrates another important aspect of genotype-phenotype correlations: they can help to identify a subset of patients for whom a certain genetic test should be performed. A correctly defined test population ensures a relatively high yield of testing, which makes testing more cost-effective and sensitive. Until recently, CANVAS was probably underdiagnosed because the diagnosis was mostly suspected in patients with the full-blown phenotype. With a better understanding of the genotype-phenotype correlation, clinicians can be made aware of the need for testing specific patient populations.

Finally, in chapter 6, we describe the rare but distinctive phenotype of sensory ataxia with variable pyramidal involvement in patients with a *RNF170* pathogenic variant. Awareness of this phenotype is important because it adds *RNF170* mutations to the relatively narrow list of differential diagnoses in patients presenting with sensory ataxia. It also promotes adding *RNF170* to gene panels for ataxia (and SPG).

As previously mentioned, repeat expansions are not yet routinely identified from MPS data in a diagnostic setting. However, progress in this area has been made and tools to detect repeat expansions have been developed, even in short-read sequencing data, although long-read sequencing is probably more reliable.[81, 112] These advancements provide a prospect on a ‘one test for all’ in the near future, where it will be possible to determine both structural variation such as repeat expansions and sequence variants from MPS data.

For ALS, the exact yield of MPS is more difficult to establish. As explained in the introduction, the non-Mendelian inheritance in most patients makes it difficult to interpret the significance of encountered pathogenic and likely pathogenic variants, even in genes previously associated with ALS. Heterogeneous methods are used to assess variant pathogenicity in sequencing studies and comparison between them is difficult. The fact that many of the potentially implicated variants are also present in controls (since they are merely risk factors) only adds to the complexity of this area. It also means that very large sample sizes are needed in order to reach sufficient statistical power to prove that some variants are truly enriched in ALS patients. Furthermore, few studies have been performed in exclusively sALS patients (Table 2). Therefore, little is known about what we can expect to find in this patient population, how we should interpret these findings and how we can counsel individual patients after MPS analysis.

	Number of patients	Sequencing technique	Number of genes in gene panel	Variants identified	Yield
Gibson et al, 2017.[113]	87 cases	WES	31	- nonsynonymous variants with European MAF <0.001 and deleterious according to MetaSVM - *C9orf72* and *ATXN2* (RL 29-33) repeat expansions	17.2%
Couthouis et al, 2014.[91]	242 cases 129 controls	WES	169	-novel and rare variants-*C9orf72* repeat expansions	cases: 60.3% controls:53.5%

Table 2: sequencing studies in exclusively sALS populations.
Repeat expansion lengths were determined separately by PCR in these studies.

Abbreviations: MAF, minor allele frequency; RL, repeat length; WES, whole exome sequencing.

With the results of chapter 3, we wanted to provide a first step in unraveling the genetic architecture of sporadic ALS by creating an inventory of the main genetic findings in a large cohort of sporadic ALS patients. We used custom ACMG-criteria which we applied in an automated way to our dataset, to identify likely pathogenic and pathogenic variants. Although we found an enrichment of these variants in cases compared to controls, the effect was still mainly driven by variants in the more strongly associated ALS-genes typically harboring variants with high effect sizes and penetrance (e.g. *SOD1*, *FUS, TARDBP*, *C9orf72*). For a rather large proportion of patients we found no clear genetic clues. This suggests that other, yet unknown factors are important for the risk in these patients. At least some of these factors are expected to be genetic, given the rather large heritability in ALS.[19] A possible strategy to identify them is to include class 3 variants in the analysis, since it can be expected that some of these variants are involved as variants with lower effect size and penetrance. However, discriminating true signal from noise is more difficult for this kind of variants and even larger sample sizes are needed, which is not easy to accomplish given the frequency of the disease. Alternatively, other variant classification systems focused on identifying variants in complex genetic diseases could be implemented.[114] It is possible that we evolve to counseling based on polygenic risk scores using MPS data, similar to what is being developed for some more common diseases with SNPs based on GWAS data.[115] One of the caveats there is that penetrance at the population level seems to be lower than what is expected based on smaller clinical cohorts, as we illustrate in our work and what has been shown for other disorders as well (e.g. CAG repeats in hereditary ataxia).[116-118]

Another possible strategy to unravel the genetics in ALS, is to look for new genes associated with the disease. In the past, the application of GWAS studies has yielded some success in this area (e.g. association with 9p21.2 (later *C9orf72* repeat expansions were identified) and the identification of *C21orf2*).[31, 119] However, as stated in the introduction, GWAS studies could only explain a small part of the estimated SNP-based heritability. It is now thought that rare variants play an

important role, but they are not easily captured by GWAS studies.[31] At this moment, some GWAS studies are still being performed but it seems that presently existing large ALS-cohorts have used most of their GWAS potential.[120] The most recent GWAS study of project MinE could only identify 15 genome-wide significant risk loci for ALS, of which eight had already been reported and two were in known genes.[121] Therefore the focus has nowadays shifted to MPS studies.
Pathway-analysis is another possible strategy to identify new genes. It has been performed for RNA-sequencing data and lately on a list of experimentally validated ALS-associated proteins.[122, 123] Recently, Saez-Atienzar et al used a polygenic risk score approach on genome-wide data to identify biological pathways in ALS.[124] However, the primary purpose of these studies was not to identify new genes, and candidates that rose from these analyses should still be confirmed. An important limitation is these study designs is again the lack of power. Common variants can be used but are known to contribute less to the disease, and for rare variant analysis large samples are needed.[124]

Genotype-phenotype correlations have been established in ALS both at gene level (e.g. frequent association of FTD and ALS in *C9orf72* repeat expansion carriers) and variant level (e.g. more aggressive disease in carriers of the *SOD1* A4V variant). However, most variables explaining disease variability remain unknown. In our study, we did not specifically look for genotype-phenotype correlations in ALS.

What with unexplained patients?

Both in our ataxia/SPG and in our ALS cohort, many patients remain genetically unexplained. Different possible explanations could underlie this finding:

1. The etiology in some patients is truly non-genetic (phenocopies).

For ataxia and SPG, many acquired etiologies can cause similar phenotypes. Although most known causes are excluded during the diagnostic work-up, we cannot eliminate the possibility that some (especially sporadic) patients actually have a non-genetic, maybe yet unknown etiology underlying their disease.
Also for ALS, acquired causes of similar phenotypes are excluded as much as possible during work-up (e.g. neuro-imaging). Many environmental factors have been suggested to be associated with the risk for ALS, but until now only for smoking the evidence is more or less convincing.[125] The marked increase in incidence of ALS in Guam around 1950 is often used as an argument for a strong environmental component of the disease. However, both the phenotype (often with marked parkinsonism and dementia) as the neuropathological findings (tauopathy) are not characteristic for the traditional form of the disease.

2. Failure to identify pathogenic variant(s) due to technical reasons.

As mentioned in the introduction, GC-rich regions, repetitive regions and detection of copy number variation (CNV) and other structural variation are still challenging for most currently used MPS-techniques. Especially for WES, coverage and read depth are not uniformly spread across the sequenced parts of the genome which means that there can be 'holes' or missing pieces in parts of the genome. As a consequence, true pathogenic variation can be missed.

3. Failure to identify pathogenic variant(s) due to lack of current knowledge.

Both for ALS and ataxia/SPG, new genes and variants are still being discovered.[47, 126] Re-analysis of MPS data of unexplained patients years after the original analysis can increase diagnostic yields, although to a limited extent.[127] The new pathogenic variants are mostly identified due to improved bioinformatics tools, updated databases and expanded knowledge on genotype-phenotype correlations. In chapter 5, we illustrate that is worthwhile to re-investigate unsolved cases when new insights become available. We screened a cohort of unexplained ataxia patients for the newly discovered *RFC1* repeat expansions with a yield of 10.6%. As explained in chapter 4, for some of these patients MPS had already been performed and it looked like the (genetic) diagnostic odyssey had come to an unfulfilling end. However, due to the analysis of *RFC1,* we have been able to provide them with a genetic diagnosis.

This being said, we suspect that most frequently prevalent variants and genes have currently been identified. Many recently discovered genes are rare causes of the disease and account for only small percentages of unexplained patients, and the question is whether all these small pieces are enough to explain the remaining missing heritability.

4. Non-coding pathogenic genetic DNA-variation.

At this moment, very little is known about the interpretation of variants in the non-coding regions of the genome. Using TS and WES, non-coding regions are not covered and can evidently not be analyzed. This is the situation for sequencing performed in the ataxia/SPG project (chapter 3). With WGS we can obtain data about the whole genome, although for many intronic regions the function is not known and possibly even obsolete. Some intronic regions are important for gene expression and regulation (e.g. untranslated regions (UTR's), promotors, enhancers and other cis-regulatory elements) or for non-coding RNAs but it is difficult to exactly delineate them.[128] This makes the interpretation of variants in intronic regions particularly challenging, although algorithms and computational models are currently developed to predict the impact of non-coding variants.[129] Despite these limitations, variants in intronic regions have been described to influence splicing and gene expression and therefore be disease-causing, also in ataxia, SPG and ALS.[130, 131] We used WGS in our ALS cohort (chapter 1) but we only analyzed the genomic regions corresponding to ALS-associated genes (based on Ensembl) and a buffer region of 2500 base pairs both upstream and downstream. In the buffer regions, we did not find any pathogenic variants in cases nor controls.

This was something we expected, because our (modified) ACMG-criteria are not suited to detect pathogenic variations in non-coding regions. One possible method to identify more of these intronic variants is a combined genome and transcriptome (RNA sequencing) approach, as has been illustrated for a cohort of patients with rare diseases.[132]

5. Other genetic mechanisms.

Not all heritable pathologies can be identified by sequencing nuclear DNA. Other mechanisms that are usually less or not detectable by MPS are pathogenic variants in mitochondrial DNA, epigenetic variation, disorders caused by imprinting problems and/or uniparental disomy and epistatic interactions. For complex genetic disorders, we expect that variants, and especially combinations of variants, in the nuclear DNA play an important role. As explained before, they are much more difficult to identify with current bioinformatics tools and knowledge.

Complex genetic diseases

In complex genetic disorders, a combination of genetic and/or environmental factors constitutes the risk for a certain disease. These disorders typically do not follow the standard Mendelian patterns of inheritance. Identified genetic factors represent mainly risk factors and additional 'hits' (either genetic or environmental) are required to develop the disease. Genetic counseling is very difficult, since many of the contributing factors remain unknown and the significance of (combinations) of genetic risk variants encountered in individual patients is hard to interpret.

ALS is considered a complex genetic disorder. However, since around 10% of the patients have a familial history very suggestive for Mendelian inheritance, we should probably look at ALS as being a spectrum as explained in chapter 1. In our study (chapter 3), we found pathogenic and likely pathogenic variants in only a proportion of sALS patients. Although we used customized ACMG-criteria, we were still biased to identify mainly rather deleterious variants (e.g. nonsense or frameshift variants) with relatively large effect size. The variants that we identified are thus probably more positioned towards the 'Mendelian' side of the spectrum. However, even such variants were present in our control cohort. This could mean that these variants are not associated with the disease, or they could identify controls with high risk on developing ALS in the future, which is not surprising given the rather high lifetime-risk of ALS. The lifetime risk is about 1/400 and thus large control populations will harbor individuals who will develop ALS at a later moment in time. An alternative explanation for the occurrence of pathogenic and likely pathogenic variants in controls could be that exhibiting such variant alone is not enough to develop the disease, consistent with an oligogenic/complex inheritance model. Supportive for this hypothesis is that we found pathogenic variants (even in *SOD1*) in a higher proportion of controls than expected. Thus, the penetrance in the general population is probably lower than deduced from smaller family-based studies and additional genetic or other factors must be present to develop the disease. Lastly, in more than 70%

of sALS patients we did not find any genetic clues. Possible reasons are listed above, but also this finding can be compatible with the oligogenic inheritance model. Variants with low effect size/penetrance are not reliably identified through our workflow and it is possible that a combination of such variants constitutes the risk on ALS. We can expect that identifying such variants in the future will be difficult, with a need for even larger sample sizes (of both cases and controls) and optimized algorithms to detect genetic variation in diseases following an oligogenic inheritance model.

Hereditary ataxia and SPG are generally considered to be Mendelian disorders. Many genes have already been linked to these diseases, in which pathogenic variants indeed follow the classical inheritance patterns.[47] However, Ngo et al recently suggested that a diagnostic rate of approximately 50% marks a relative ceiling for current WES-based methods applied for ataxia and related neurodegenerative diseases.[133] Although missed CNVs, repeat expansions, other structural variation, noncoding variants and variants in yet unknown genes undoubtedly account for at least a part of the unexplained cases, the question remains if these will be enough to close the diagnostic gap of currently unexplained patients.[134] Hypothetically, it is possible that some patients display a non-Mendelian inheritance. As is the case for ALS, ataxia and SPG should then be regarded as a spectrum between Mendelian and more complex inheritance models, though relatively more weight should be given to Mendelian inheritance compared to ALS since more patients display a familial history. Especially for the group of adult-onset sporadic patients, this model could explain the relatively low yields obtained with MPS-based methods although another plausible explanation are non-genetic etiologies in this group. The reduced penetrance (or higher prevalence) of polyglutamine repeat expansions of some SCA's among large population based cohorts compared to the expected rates based on smaller clinical cohorts, is another argument that other, secondary (genetic) factors might be needed for the disease to be penetrant.[117] In chapter 4, we looked for heterozygous variants in genes associated with AR ataxia and SPG.

Some carriers of such heterozygous variants can probably be explained by a second hit in the same gene, currently not yet identifiable by (short read) exome sequencing. It has been shown that for some patients a second variant can be identified by implementing additional techniques. An example is given by Verdura et al.[130] They describe a patient with a missense variant in *SPG7* identified by WES, for whom a second, deep intronic splice variant was found by performing WGS (and reverse transcriptase-PCR (RT-PCR) to confirm that the variant activated a cryptic splice site). With long read sequencing, other variants such as structural variants (e.g. larger insertions, deletions, duplications, inversions or translocations with or without gene fusions) can be established; repetitive regions can more reliably be covered and base modifications can be analyzed.[135] This technique could thus potentially identify second hits for the patients carrying a heterozygous variant in AR ataxia and SPG genes. However, looking at the patients carrying these variants in our cohort, their

phenotype often does not fit the phenotype typically associated with the genes carrying the heterozygous variants (data not shown). Therefore, we hypothesize that such variants might not be enough to cause disease on their own and that a second hit in the same gene is rather unlikely given the phenotype of most patients. However, the heterozygous variants could be pathogenic together with some other genetic risk factors, which is consistent with an oligogenic inheritance model. It has to be said that these results are only descriptive, especially since we did not compare the frequency of these variants in cases with the frequency in a control population. As for ALS, further research on large cohorts is needed to confirm the presence of such risk variants and the implication of combinations of variants in ataxia and SPG.

In any case, for all these disorders a great challenge lies ahead to further unravel the underlying genetic etiologies. Only with a correct understanding of the genetic architecture underlying a disease, patients can accurately be counseled about risks and the implications for themselves and their family.

Overlap and selective vulnerability of the motor system

In this thesis, we 'dissected' some disorders typically affecting the motor system. We demonstrated that overlap exists between these disorders, e.g. in chapter 4, where we explored genotypes and phenotypes of patients with hereditary ataxia and SPG, as well as in chapter 5, where we described the phenotypic and genetic spectrum of patients with *RNF170* variants. Phenotypic overlap was illustrated in patients displaying signs of involvement of the UMNs, LMNs and/or the cerebellar system, even when they carried pathogenic variants in genes typically associated with only one of the phenotypes. On the other hand, some genes were found to typically overlap between the phenotypes, e.g. *SPG7*, in which variants can cause SPG, ataxia or mixed spastic ataxia. In *RNF170*, the main phenotype is even dependent on the mutation type and inheritance mode. The 3D plot in chapter 4 illustrates that variants in many genes can affect more than one system. Some genes show higher affinities with certain parts of the motor system, but a lot of variability exists. All this underscores the potential overlap between the different phenotypes, as has also been suggested by others.[5, 134, 136, 137] However, if this overlap is an exceptional event or the first glance of a shared genetic architecture, remains unexplored. We hypothesize that all parts of the motor system exhibit a shared genetic vulnerability, probably rooted in shared molecular pathways. On the other hand, each part of the motor system shows unique characteristics making that each has their own specific vulnerabilities as well. This could explain why variants in some genes are more prone to result in specific phenotypes.

The advent of next generation sequencing technologies allows for a comprehensive analysis of genetic causes in patients with neurodegenerative disorders affecting primarily or in part the motor system. By studying these disorders, we can identify important genetic determinants of motor neuron

vulnerability and improve our understanding of mechanisms of neurodegenerative disorders. In this thesis, we aimed to gain more insight in the genetic determinants of these specific disorders. It would be interesting, as a next step, to use a novel multi-disease approach. This would comprise pooling all disorders together and look for genetic determinants of the different parts of the motor system (e.g. pyramidal, peripheral or cerebellar system) using bio-informatic modeling. However, large datasets are needed for this purpose, which is not straightforward for this group of rare diseases.

Future perspectives

As can be appreciated from what is written above, many questions remain about the genetic architecture of motor neuron disorders. Further knowledge is needed, both about the disorders as separate entities as well as about the general vulnerability of the motor system. A better understanding of the genetic determinants of motor neuron vulnerability is indispensable for developing novel therapeutic strategies for these severe incurable disorders.

Concerning the distinct parts of the motor system, it would be interesting to also genetically dissect the disorders affecting mainly the peripheral system, being the inherited neuropathies. This would mean characterizing a cohort of patients with Charcot-Marie-Tooth disease (CMT), or more specifically distal hereditary motor neuropathy (dHMN), and map the spectrum of genetic etiologies encountered in Belgium. MPS techniques could be used to increase the yield in currently unexplained patients, although at first glimpse, the yield of gene panel analysis for dHMN in our center seems to be lower compared to the yield for ataxia and SPG (data not shown). However, for many patients the analysis dates back from several years ago and recent studies have shown higher yields of around 30%.[138, 139]

The existence of unexplained patients with disorders affecting the motor system, especially the ones with a familial history, opens possibilities to look for novel disease-causing gene mutations. These pedigrees are often too small to calculate LOD-scores, but WGS allows segregation analysis in smaller families. Subsequently, candidate variants and genes can be compared to other patients harboring the same phenotype, both in-house as by the use of sharing platforms as Matchmaker Exhange.[140]

Lastly, as mentioned before, it would be very interesting to study the vulnerability of the motor system using a multi-disease approach and bioinformatics modeling. Ideally, large cohorts with genetic and phenotypic data should therefore be available. For rare diseases as ataxia, SPG and ALS, analysis in the context of large international consortia would be preferable.

Thesis abstract

BACKGROUND: Amyotrophic lateral sclerosis (ALS), hereditary ataxia (HA) and hereditary spastic paraplegia (HSP) are disorders affecting the motor system. They all have an important genetic basis and are characterized by genetic and phenotypic heterogeneity. For ALS, a minority of patients carry rare variants with high penetrance and display a Mendelian inheritance pattern. However, most patients are sporadic and are thought to exhibit an oligogenic inheritance. HA and HSP are considered Mendelian disorders, but many patients remain unexplained despite thorough (genetic) investigations. The phenotypic, genetic and mechanistic overlap between the three disorders arises suspicion for a shared genetic vulnerability of the motor system and places the involved phenotypes along a continuum. The advent of next generation sequencing (MPS) technologies opens opportunities to dissect this shared genetic architecture and explore genetic determinants of motor neuron vulnerability.

OBJECTIVE AND METHODS: The objective of this PhD project was to gain more insight in the genetic determinants of disorders affecting the motor system. Therefore, we mapped pathogenic mutations within each phenotypes and established gene frequencies in Belgium (for HA and HSP), we explored genotype-phenotype correlations and looked for genetic and phenotypic overlap between the different disorders affecting the motor system. We performed retrospective chart analysis and used MPS analysis.

RESULTS: For ALS, we provided an inventory of pathogenic and likely pathogenic genetic variation in a large cohort of sporadic ALS patients using customized ACMG-criteria. We found a significant enrichment of pathogenic variants in cases, although for more than 70% of patients, we could not find genetic clues. Also not for all ALS-associated genes evidence for association could be provided. In line with the oligogenic hypothesis, we found significantly more combinations of variants in cases compared to controls. For HA and HSP, we determined the yield of targeted gene testing (36.2%) and subsequent gene panel analysis (28.9%) and established gene frequencies in Belgium. We described genotype-phenotype correlations and defined them in more detail for a family with a pathogenic *RNF170* variant, thereby also evaluating phenotypic overlap. We explored the yield of *RFC1* repeat expansion analysis in unexplained ataxia patients (10.6%) and established that patients with a concomitant (subclinical) neuropathy are the best target population. Contemplating non-Mendelian inheritance in a subset of unexplained patients with ataxia and/or spastic paraplegia, we listed heterozygous variants in genes linked with autosomal recessive HA or HSP. We established genetic and clinical overlap between the phenotypes.

CONCLUSION: Our data underwrite the idea of a shared genetic vulnerability of the motor system. However, many questions remain about the genetic architecture of disorders affecting the motor system. A better understanding is important for genetic counseling and for the development of novel therapeutic strategies for these severe incurable diseases. Our work can serve as a valuable resource for counseling and for the assembly and analysis of gene panels for these disorders.

References related to introduction and discussion

1. Bennett MR, Hacker PM. The motor system in neuroscience: a history and analysis of conceptual developments. *Prog Neurobiol* 2002;67(1):1-52.
2. Blumenfeld H. Neuroanatomy through clinical cases. Second edition ed: Sinauer Associates; 2010.
3. Hardiman O, Al-Chalabi A, Chio A, et al. Amyotrophic lateral sclerosis. *Nat Rev Dis Primers* 2017;3:17085.
4. Tesson C, Koht J, Stevanin G. Delving into the complexity of hereditary spastic paraplegias: how unexpected phenotypes and inheritance modes are revolutionizing their nosology. *Hum Genet* 2015;134(6):511-38.
5. Synofzik M, Schule R. Overcoming the divide between ataxias and spastic paraplegias: Shared phenotypes, genes, and pathways. *Mov Disord* 2017;32(3):332-45.
6. Alonso A, Logroscino G, Jick SS, et al. Incidence and lifetime risk of motor neuron disease in the United Kingdom: a population-based study. *Eur J Neurol* 2009;16(6):745-51.
7. GBD 2016 Motor Neuron Disease Collaborators. Global, regional, and national burden of motor neuron diseases 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol* 2018;17(12):1083-97.
8. Volk AE, Weishaupt JH, Andersen PM, et al. Current knowledge and recent insights into the genetic basis of amyotrophic lateral sclerosis. *Med Genet* 2018;30(2):252-8.
9. Bensimon G, Lacomblez L, Meininger V. A controlled trial of riluzole in amyotrophic lateral sclerosis. ALS/Riluzole Study Group. *N Engl J Med* 1994;330(9):585-91.
10. Palumbo JM, Hubble J, Apple S, et al. Post-hoc analyses of the edaravone clinical trials Study 16 and Study 19: a step toward more efficient clinical trial designs in amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Frontotemporal Degener* 2019;20(5-6):421-31.
11. Klim JR, Vance C, Scotter EL. Antisense oligonucleotide therapies for Amyotrophic Lateral Sclerosis: Existing and emerging targets. *Int J Biochem Cell Biol* 2019;110:149-53.
12. Amado DA, Davidson BL. Gene therapy for ALS: A review. *Mol Ther* 2021;29(12):3345-58.
13. Byrne S, Heverin M, Elamin M, et al. Aggregation of neurologic and neuropsychiatric disease in amyotrophic lateral sclerosis kindreds: a population-based case-control cohort study of familial and sporadic amyotrophic lateral sclerosis. *Ann Neurol* 2013;74(5):699-708.
14. Byrne S, Bede P, Elamin M, et al. Proposed criteria for familial amyotrophic lateral sclerosis. *Amyotroph Lateral Scler* 2011;12(3):157-9.
15. Leblond CS, Kaneb HM, Dion PA, et al. Dissection of genetic factors associated with amyotrophic lateral sclerosis. *Exp Neurol* 2014;262 Pt B:91-101.
16. Renton AE, Chio A, Traynor BJ. State of play in amyotrophic lateral sclerosis genetics. *Nat Neurosci* 2014;17(1):17-23.
17. Renton AE, Majounie E, Waite A, et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* 2011;72(2):257-68.
18. Van Damme P, Robberecht W, Van Den Bosch L. Modelling amyotrophic lateral sclerosis: progress and possibilities. *Dis Model Mech* 2017;10(5):537-49.
19. Al-Chalabi A, Fang F, Hanby MF, et al. An estimate of amyotrophic lateral sclerosis heritability using twin data. *J Neurol Neurosurg Psychiatry* 2010;81(12):1324-6.
20. Hanby MF, Scott KM, Scotton W, et al. The risk to relatives of patients with sporadic amyotrophic lateral sclerosis. *Brain* 2011;134(Pt 12):3454-7.
21. Chio A, Battistini S, Calvo A, et al. Genetic counselling in ALS: facts, uncertainties and clinical suggestions. *J Neurol Neurosurg Psychiatry* 2014;85(5):478-85.
22. van Blitterswijk M, van Es MA, Hennekam EA, et al. Evidence for an oligogenic basis of amyotrophic lateral sclerosis. *Hum Mol Genet* 2012;21(17):3776-84.
23. He J, Mangelsdorf M, Fan D, et al. Amyotrophic Lateral Sclerosis Genetic Studies: From Genome-wide Association Mapping to Genome Sequencing. *Neuroscientist* 2015;21(6):599-615.

24. Al-Chalabi A, Calvo A, Chio A, et al. Analysis of amyotrophic lateral sclerosis as a multistep process: a population-based modelling study. *Lancet Neurol* 2014;13(11):1108-13.
25. Tazelaar GHP, Boeynaems S, De Decker M, et al. ATXN1 repeat expansions confer risk for amyotrophic lateral sclerosis and contribute to TDP-43 mislocalization. *Brain Commun* 2020;2(2):fcaa064.
26. Elden AC, Kim HJ, Hart MP, et al. Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature* 2010;466(7310):1069-75.
27. Blauw HM, van Rheenen W, Koppers M, et al. NIPA1 polyalanine repeat expansions are associated with amyotrophic lateral sclerosis. *Hum Mol Genet* 2012;21(11):2497-502.
28. van der Zee J, Gijselinck I, Van Mossevelde S, et al. TBK1 Mutation Spectrum in an Extended European Patient Cohort with Frontotemporal Dementia and Amyotrophic Lateral Sclerosis. *Hum Mutat* 2017;38(3):297-309.
29. Kenna KP, van Doormaal PT, Dekker AM, et al. NEK1 variants confer susceptibility to amyotrophic lateral sclerosis. *Nat Genet* 2016;48(9):1037-42.
30. van Es MA, Veldink JH, Saris CG, et al. Genome-wide association study identifies 19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. *Nat Genet* 2009;41(10):1083-7.
31. van Rheenen W, Shatunov A, Dekker AM, et al. Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. *Nat Genet* 2016;48(9):1043-8.
32. Van Damme P. How much of the missing heritability of ALS is hidden in known ALS genes? *J Neurol Neurosurg Psychiatry* 2018;89(8):794.
33. Blauw HM, Al-Chalabi A, Andersen PM, et al. A large genome scan for rare CNVs in amyotrophic lateral sclerosis. *Hum Mol Genet* 2010;19(20):4091-9.
34. Lattante S, Pomponi MG, Conte A, et al. ATXN1 intermediate-length polyglutamine expansions are associated with amyotrophic lateral sclerosis. *Neurobiol Aging* 2018;64:157 e1- e5.
35. Diekstra FP, Van Deerlin VM, van Swieten JC, et al. C9orf72 and UNC13A are shared risk loci for amyotrophic lateral sclerosis and frontotemporal dementia: a genome-wide meta-analysis. *Ann Neurol* 2014;76(1):120-33.
36. Dewan R, Chia R, Ding J, et al. Pathogenic Huntingtin Repeat Expansions in Patients with Frontotemporal Dementia and Amyotrophic Lateral Sclerosis. *Neuron* 2021;109(3):448-60 e4.
37. Campbell W. DeJong's The Neurologic Examination. 7th ed: Lippincott Williams & Wilkins; 2013.
38. Chhetri SK, Gow D, Shaunak S, et al. Clinical assessment of the sensory ataxias; diagnostic algorithm with illustrative cases. *Pract Neurol* 2014;14(4):242-51.
39. van Gaalen J, van de Warrenburg BP. A practical approach to late-onset cerebellar ataxia: putting the disorder with lack of order into order. *Pract Neurol* 2012;12(1):14-24.
40. Barsottini OG, Albuquerque MV, Braga-Neto P, et al. Adult onset sporadic ataxias: a diagnostic challenge. *Arq Neuropsiquiatr* 2014;72(3):232-40.
41. Abele M, Burk K, Schols L, et al. The aetiology of sporadic adult-onset ataxia. *Brain* 2002;125(5):961-8.
42. Anheim M, Fleury M, Monga B, et al. Epidemiological, clinical, paraclinical and molecular study of a cohort of 102 patients affected with autosomal recessive progressive cerebellar ataxia from Alsace, Eastern France: implications for clinical management. *Neurogenetics* 2010;11(1):1-12.
43. Klockgether T. Sporadic ataxia with adult onset: classification and diagnostic criteria. *Lancet Neurol* 2010;9(1):94-104.
44. Jayadev S, Bird TD. Hereditary ataxias: overview. *Genet Med* 2013;15(9):673-83.
45. Geschwind DH, Paulson H, Klein C. Neurogenetics Part I. Aminoff M, Boller F, Swaab D, editors: Elsevier; 2018.
46. Klockgether T, Mariotti C, Paulson HL. Spinocerebellar ataxia. *Nature Reviews Disease Primers* 2019;5.
47. Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD) [Internet]. Available from: https://omim.org/.

48. Nussbaum RL, McInnes RR, Willard HF. Thompson & Thompson Genetics in medicine. Eight edition ed: Elsevier; 2016.
49. Read A, Donnai D. New clinical genetics, a guide to genomic medicine. Fourth edition ed: Scion PublishingLimited; 2021.
50. Cortese A, Simone R, Sullivan R, et al. Biallelic expansion of an intronic repeat in RFC1 is a common cause of late-onset ataxia. *Nat Genet* 2019;51(4):649-58.
51. Cortese A, Tozza S, Yau WY, et al. Cerebellar ataxia, neuropathy, vestibular areflexia syndrome due to RFC1 repeat expansion. *Brain* 2020;143(2):480-90.
52. Storey E. Genetic cerebellar ataxias. *Semin Neurol* 2014;34(3):280-92.
53. Nibbeling EAR, Duarri A, Verschuuren-Bemelmans CC, et al. Exome sequencing and network analysis identifies shared mechanisms underlying spinocerebellar ataxia. *Brain* 2017;140(11):2860-78.
54. Coutelier M, Hammer MB, Stevanin G, et al. Efficacy of Exome-Targeted Capture Sequencing to Detect Mutations in Known Cerebellar Ataxia Genes. *JAMA Neurol* 2018;75(5):591-9.
55. Ocana-Santero G, Diaz-Nido J, Herranz-Martin S. Future Prospects of Gene Therapy for Friedreich's Ataxia. *Int J Mol Sci* 2021;22(4).
56. Traschutz A, Reich S, Adarmes AD, et al. The ARCA Registry: A Collaborative Global Platform for Advancing Trial Readiness in Autosomal Recessive Cerebellar Ataxias. *Front Neurol* 2021;12:677551.
57. Fink JK. Hereditary spastic paraplegia: clinical principles and genetic advances. *Semin Neurol* 2014;34(3):293-305.
58. Klebe S, Stevanin G, Depienne C. Clinical and genetic heterogeneity in hereditary spastic paraplegias: from SPG1 to SPG72 and still counting. *Rev Neurol (Paris)* 2015;171(6-7):505-30.
59. Hereditary Spastic Paraplegia Overview. 2000 Aug 15 [Updated 2021 Feb 11]. In: Adam MP, Ardinger HH, Pagon RA, et al., editors. GeneReviews® [Internet] [Internet]. Seattle (WA): University of Washington, Seattle. 1993-2021. Available from: https://www.ncbi.nlm.nih.gov/books/NBK1509/. Available from: https://www.ncbi.nlm.nih.gov/books/NBK1509/.
60. Lo Giudice T, Lombardi F, Santorelli FM, et al. Hereditary spastic paraplegia: clinical-genetic characteristics and evolving molecular mechanisms. *Exp Neurol* 2014;261:518-39.
61. Lallemant-Dudek P, Darios F, Durr A. Recent advances in understanding hereditary spastic paraplegias and emerging therapies. *Fac Rev* 2021;10:27.
62. Huin V, Coarelli G, Guemy C, et al. Motor neuron pathology in CANVAS due to RFC1 expansions. *Brain* 2021.
63. Citterio A, Arnoldi A, Panzeri E, et al. Variants in KIF1A gene in dominant and sporadic forms of hereditary spastic paraparesis. *J Neurol* 2015;262(12):2684-90.
64. Delva A, Thakore N, Pioro EP, et al. Finger extension weakness and downbeat nystagmus motor neuron disease syndrome: A novel motor neuron disorder? *Muscle Nerve* 2017;56(6):1164-8.
65. van de Warrenburg BP, Schouten MI, de Bot ST, et al. Clinical exome sequencing for cerebellar ataxia and spastic paraplegia uncovers novel gene-disease associations and unanticipated rare disorders. *Eur J Hum Genet* 2016;24(10):1460-6.
66. Casari G, De Fusco M, Ciarmatori S, et al. Spastic paraplegia and OXPHOS impairment caused by mutations in paraplegin, a nuclear-encoded mitochondrial metalloprotease. *Cell* 1998;93(6):973-83.
67. Pfeffer G, Pyle A, Griffin H, et al. SPG7 mutations are a common cause of undiagnosed ataxia. *Neurology* 2015;84(11):1174-6.
68. Qiu Y, Zhong S, Cong L, et al. A novel KIF5A gene variant causes spastic paraplegia and cerebellar ataxia. *Ann Clin Transl Neurol* 2018;5(11):1415-20.
69. Van Damme P, Veldink JH, van Blitterswijk M, et al. Expanded ATXN2 CAG repeat size in ALS identifies genetic overlap between ALS and SCA2. *Neurology* 2011;76(24):2066-72.
70. Moreira MC, Klur S, Watanabe M, et al. Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2. *Nat Genet* 2004;36(3):225-7.
71. Chen YZ, Bennett CL, Huynh HM, et al. DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). *Am J Hum Genet* 2004;74(6):1128-35.

72. Vallat JM, Goizet C, Tazir M, et al. Classifications of neurogenetic diseases: An increasingly complex problem. *Rev Neurol (Paris)* 2016;172(6-7):339-49.
73. Orlacchio A, Babalini C, Borreca A, et al. SPATACSIN mutations cause autosomal recessive juvenile amyotrophic lateral sclerosis. *Brain* 2010;133(Pt 2):591-8.
74. Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet* 2010;11(1):31-46.
75. Galatolo D, Tessa A, Filla A, et al. Clinical application of next generation sequencing in hereditary spinocerebellar ataxia: increasing the diagnostic yield and broadening the ataxia-spasticity spectrum. A retrospective analysis. *Neurogenetics* 2018;19(1):1-8.
76. Head SR, Komori HK, LaMere SA, et al. Library construction for next-generation sequencing: overviews and challenges. *Biotechniques* 2014;56(2):61-4, 6, 8, passim.
77. Bamshad MJ, Ng SB, Bigham AW, et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* 2011;12(11):745-55.
78. Meienberg J, Bruggmann R, Oexle K, et al. Clinical sequencing: is WGS the better WES? *Hum Genet* 2016;135(3):359-62.
79. Gorcenco S, Ilinca A, Almasoudi W, et al. New generation genetic testing entering the clinic. *Parkinsonism Relat Disord* 2020;73:72-84.
80. Dolzhenko E, Deshpande V, Schlesinger F, et al. ExpansionHunter: a sequence-graph-based tool to analyze variation in short tandem repeat regions. *Bioinformatics* 2019;35(22):4754-6.
81. Dolzhenko E, Bennett MF, Richmond PA, et al. ExpansionHunter Denovo: a computational method for locating known and novel repeat expansions in short-read sequencing data. *Genome Biol* 2020;21(1):102.
82. Pollard MO, Gurdasani D, Mentzer AJ, et al. Long reads: their purpose and place. *Hum Mol Genet* 2018;27(R2):R234-R41.
83. Muller K, Brenner D, Weydt P, et al. Comprehensive analysis of the mutation spectrum in 301 German ALS families. *J Neurol Neurosurg Psychiatry* 2018;89(8):817-27.
84. Nishiyama A, Niihori T, Warita H, et al. Comprehensive targeted next-generation sequencing in Japanese familial amyotrophic lateral sclerosis. *Neurobiol Aging* 2017;53:194 e1- e8.
85. Morgan S, Shatunov A, Sproviero W, et al. A comprehensive analysis of rare genetic variation in amyotrophic lateral sclerosis in the UK. *Brain* 2017;140(6):1611-8.
86. Narain P, Pandey A, Gupta S, et al. Targeted next-generation sequencing reveals novel and rare variants in Indian patients with amyotrophic lateral sclerosis. *Neurobiol Aging* 2018;71:265 e9-e14.
87. Gratten J, Zhao Q, Benyamin B, et al. Whole-exome sequencing in amyotrophic lateral sclerosis suggests NEK1 is a risk gene in Chinese. *Genome Med* 2017;9(1):97.
88. Cirulli ET, Lasseigne BN, Petrovski S, et al. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science* 2015;347(6229):1436-41.
89. Pang SY, Hsu JS, Teo KC, et al. Burden of rare variants in ALS genes influences survival in familial and sporadic ALS. *Neurobiol Aging* 2017;58:238 e9- e15.
90. Kenna KP, McLaughlin RL, Byrne S, et al. Delineating the genetic heterogeneity of ALS using targeted high-throughput sequencing. *J Med Genet* 2013;50(11):776-83.
91. Couthouis J, Raphael AR, Daneshjou R, et al. Targeted exon capture and sequencing in sporadic amyotrophic lateral sclerosis. *PLoS Genet* 2014;10(10):e1004704.
92. Richards S, Aziz N, Bale 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* 2015;17(5):405-24.
93. Debray S, Race V, Crabbe V, et al. Frequency of C9orf72 repeat expansions in amyotrophic lateral sclerosis: a Belgian cohort study. *Neurobiol Aging* 2013;34(12):2890 e7- e12.
94. Pyle A, Smertenko T, Bargiela D, et al. Exome sequencing in undiagnosed inherited and sporadic ataxias. *Brain* 2015;138(Pt 2):276-83.
95. Nemeth AH, Kwasniewska AC, Lise S, et al. Next generation sequencing for molecular diagnosis of neurological disorders using ataxias as a model. *Brain* 2013;136(Pt 10):3106-18.
96. Fogel BL, Lee H, Deignan JL, et al. Exome sequencing in the clinical diagnosis of sporadic or familial cerebellar ataxia. *JAMA Neurol* 2014;71(10):1237-46.

97. Ohba C, Osaka H, Iai M, et al. Diagnostic utility of whole exome sequencing in patients showing cerebellar and/or vermis atrophy in childhood. *Neurogenetics* 2013;14(3-4):225-32.
98. Sawyer SL, Schwartzentruber J, Beaulieu CL, et al. Exome sequencing as a diagnostic tool for pediatric-onset ataxia. *Hum Mutat* 2014;35(1):45-9.
99. Sun M, Johnson AK, Nelakuditi V, 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* 2019;21(1):195-206.
100. Coutelier M, Coarelli G, Monin ML, et al. A panel study on patients with dominant cerebellar ataxia highlights the frequency of channelopathies. *Brain* 2017;140(6):1579-94.
101. Novarino G, Fenstermaker AG, Zaki MS, et al. Exome sequencing links corticospinal motor neuron disease to common neurodegenerative disorders. *Science* 2014;343(6170):506-11.
102. Burguez D, Polese-Bonatto M, Scudeiro LAJ, et al. Clinical and molecular characterization of hereditary spastic paraplegias: A next-generation sequencing panel approach. *J Neurol Sci* 2017;383:18-25.
103. Schule R, Wiethoff S, Martus P, et al. Hereditary spastic paraplegia: Clinicogenetic lessons from 608 patients. *Ann Neurol* 2016;79(4):646-58.
104. Kara E, Tucci A, Manzoni C, et al. Genetic and phenotypic characterization of complex hereditary spastic paraplegia. *Brain* 2016;139(Pt 7):1904-18.
105. Kumar KR, Blair NF, Vandebona H, et al. Targeted next generation sequencing in SPAST-negative hereditary spastic paraplegia. *J Neurol* 2013;260(10):2516-22.
106. Lynch DS, Koutsis G, Tucci A, et al. Hereditary spastic paraplegia in Greece: characterisation of a previously unexplored population using next-generation sequencing. *Eur J Hum Genet* 2016;24(6):857-63.
107. Balicza P, Grosz Z, Gonzalez MA, et al. Genetic background of the hereditary spastic paraplegia phenotypes in Hungary - An analysis of 58 probands. *J Neurol Sci* 2016;364:116-21.
108. Project MinE ALS Sequencing Consortium. Project MinE: study design and pilot analyses of a large-scale whole-genome sequencing study in amyotrophic lateral sclerosis. *Eur J Hum Genet* 2018;26:1537-46.
109. Giordano I, Harmuth F, Jacobi H, et al. Clinical and genetic characteristics of sporadic adult-onset degenerative ataxia. *Neurology* 2017;89(10):1043-9.
110. Wan L, Chen Z, Wan N, et al. Biallelic Intronic AAGGG Expansion of RFC1 is Related to Multiple System Atrophy. *Ann Neurol* 2020;88(6):1132-43.
111. Sullivan R, Yau WY, Chelban V, et al. RFC1-related ataxia is a mimic of early multiple system atrophy. *J Neurol Neurosurg Psychiatry* 2021.
112. Mitsuhashi S, Frith MC, Mizuguchi T, et al. Tandem-genotypes: robust detection of tandem repeat expansions from long DNA reads. *Genome Biol* 2019;20(1):58.
113. Gibson SB, Downie JM, Tsetsou S, et al. The evolving genetic risk for sporadic ALS. *Neurology* 2017;89(3):226-33.
114. Lvovs D, Favorova OO, Favorov AV. A Polygenic Approach to the Study of Polygenic Diseases. *Acta Naturae* 2012;4(3):59-71.
115. Ge T, Chen CY, Ni Y, et al. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun* 2019;10(1):1776.
116. Kay C, Collins JA, Miedzybrodzka Z, et al. Huntington disease reduced penetrance alleles occur at high frequency in the general population. *Neurology* 2016;87(3):282-8.
117. Gardiner SL, Boogaard MW, Trompet S, et al. Prevalence of Carriers of Intermediate and Pathological Polyglutamine Disease-Associated Alleles Among Large Population-Based Cohorts. *JAMA Neurol* 2019;76(6):650-6.
118. Van Hout CV, Tachmazidou I, Backman JD, et al. Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature* 2020;586(7831):749-56.
119. Fogh I, Ratti A, Gellera C, et al. A genome-wide association meta-analysis identifies a novel locus at 17q11.2 associated with sporadic amyotrophic lateral sclerosis. *Hum Mol Genet* 2014;23(8):2220-31.
120. Nakamura R, Misawa K, Tohnai G, et al. A multi-ethnic meta-analysis identifies novel genes, including ACSL5, associated with amyotrophic lateral sclerosis. *Commun Biol* 2020;3(1):526.

121. van Rheenen W, van der Spek RAA, Bakker MK, et al. Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology. *Nat Genet* 2021;53(12):1636-48.
122. Wang JC, Ramaswami G, Geschwind DH. Gene co-expression network analysis in human spinal cord highlights mechanisms underlying amyotrophic lateral sclerosis susceptibility. *Sci Rep* 2021;11(1):5748.
123. Kumar R, Haider S. Protein network analysis to prioritize key genes in amyotrophic lateral sclerosis. *IBRO Neurosci Rep* 2022;12:25-44.
124. Saez-Atienzar S, Bandres-Ciga S, Langston RG, et al. Genetic analysis of amyotrophic lateral sclerosis identifies contributing pathways and cell types. *Sci Adv* 2021;7(3).
125. Oskarsson B, Horton DK, Mitsumoto H. Potential Environmental Factors in Amyotrophic Lateral Sclerosis. *Neurol Clin* 2015;33(4):877-88.
126. Al-Chalabi A, van den Berg LH, Veldink J. Gene discovery in amyotrophic lateral sclerosis: implications for clinical management. *Nat Rev Neurol* 2017;13(2):96-104.
127. Won D, Kim SH, Kim B, et al. Reanalysis of Genomic Sequencing Results in a Clinical Laboratory: Advantages and Limitations. *Front Neurol* 2020;11:612.
128. Esteller M. Non-coding RNAs in human disease. *Nat Rev Genet* 2011;12(12):861-74.
129. Lin H, Hargreaves KA, Li R, et al. RegSNPs-intron: a computational framework for predicting pathogenic impact of intronic single nucleotide variants. *Genome Biol* 2019;20(1):254.
130. Verdura E, Schluter A, Fernandez-Eulate G, et al. A deep intronic splice variant advises reexamination of presumably dominant SPG7 Cases. *Ann Clin Transl Neurol* 2020;7(1):105-11.
131. Saez-Atienzar S, Dalgard CL, Ding J, et al. Identification of a pathogenic intronic KIF5A mutation in an ALS-FTD kindred. *Neurology* 2020;95(22):1015-8.
132. Lee H, Huang AY, Wang LK, et al. Diagnostic utility of transcriptome sequencing for rare Mendelian diseases. *Genet Med* 2020;22(3):490-9.
133. Ngo KJ, Rexach JE, Lee H, et al. A diagnostic ceiling for exome sequencing in cerebellar ataxia and related neurological disorders. *Hum Mutat* 2020;41(2):487-501.
134. Bis-Brewer DM, Zuchner S. Perspectives on the Genomics of HSP Beyond Mendelian Inheritance. *Front Neurol* 2018;9:958.
135. Amarasinghe SL, Su S, Dong X, et al. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol* 2020;21(1):30.
136. Parodi L, Fenu S, Stevanin G, et al. Hereditary spastic paraplegia: More than an upper motor neuron disease. *Rev Neurol (Paris)* 2017;173(5):352-60.
137. Strong MJ, Gordon PH. Primary lateral sclerosis, hereditary spastic paraplegia and amyotrophic lateral sclerosis: discrete entities or spectrum? *Amyotroph Lateral Scler Other Motor Neuron Disord* 2005;6(1):8-16.
138. Liu X, Duan X, Zhang Y, et al. Molecular analysis and clinical diversity of distal hereditary motor neuropathy. *Eur J Neurol* 2020;27(7):1319-26.
139. Bansagi B, Griffin H, Whittaker RG, et al. Genetic heterogeneity of motor neuropathies. *Neurology* 2017;88(13):1226-34.
140. Philippakis AA, Azzariti DR, Beltran S, et al. The Matchmaker Exchange: a platform for rare disease gene discovery. *Hum Mutat* 2015;36(10):915-21.

Scientific acknowledgements, personal contributions and conflicts of interest

Chapter 3: Genetic variability in known ALS genes in sporadic ALS patients

Contributions: Study concept and design: Philip Van Damme, Sien H. Van Daele, Matthieu M. Moisse. Acquisition of data: Sien H. Van Daele, Matthieu Moisse. Analysis and interpretation of data: Sien H. Van Daele, Matthieu Moisse. Critical revision of the manuscript for important intellectual content: Philip Van Damme. Supervision: Philip Van Damme.
Acknowledgements: We thank the patients participating in this study. This project could be realized as part of the international consortium of Project MinE. SHVD is funded by a PhD fellowship of the Research Foundation - Flanders (FWO) (1164018N, file number 40900). Research is furthermore funded by the Fund for Clinical Academic studies of the university hospitals of Leuven, Belgium (S61184). PVD is Senior Clinical investigators of FWO. PVD is supported by the ALS Liga België and the KU Leuven funds “Een Hart voor ALS”, “Laeversfonds voor ALS Onderzoek” and the “Valéry Perrier Race against ALS Fund”. PVD holds the Emil von Behring Chair in Neuromuscular and Neurodegenerative Disorders by CSL Behring.
Conflicts of interest: none.

Chapter 4: Genetic yield and phenotype-genotype correlations in a large ataxia-HSP cohort in Belgium

Contributions: Study concept and design: Philip Van Damme, Sien H. Van Daele. Acquisition of data: Sien H. Van Daele. Analysis and interpretation of data: Sien H. Van Daele. Critical revision of the manuscript for important intellectual content: Philip Van Damme. Critical revision of the manuscript for important intellectual content: Philip Van Damme, Kristl G. Claeys, Sascha Vermeer, Hilde Van Esch, Wim Vandenberghe, Liesbeth De Waele, Valérie Race, Amélie Van Eesbeeck, Liesbeth Keldermans, Matthieu Moisse. Supervision: Philip Van Damme.
Acknowledgements: We thank the patients participating in this study. SHVD is funded by a PhD fellowship of the Research Foundation - Flanders (FWO) (1164018N, file number 40900). Research is furthermore funded by the Fund for Clinical Academic studies of the university hospitals of Leuven, Belgium (S61184). PVD is Senior Clinical investigators of FWO. PVD is supported by the ALS Liga België and the KU Leuven funds “Een Hart voor ALS”, “Laeversfonds voor ALS Onderzoek” and the “Valéry Perrier Race against ALS Fund”. PVD holds the Emil von Behring Chair in Neuromuscular and Neurodegenerative Disorders by CSL Behring.
Conflicts of interest: none.

Chapter 5: Diagnostic yield of testing for *RFC1* repeat expansions in patients with unexplained adult-onset cerebellar ataxia

Contributions: Study concept and design: Wim Vandenberghe. Acquisition of Data: Sien H. Van Daele, Laura Lannoo, Wim Vandenberghe, Amélie Van Eesbeeck. Analysis and interpretation of data: Sien H. Van Daele, Sascha Vermeer, Amélie Van Eesbeeck, Wim Vandenberghe. Drafting the manuscript: Sien H. Van Daele, Sascha Vermeer. Critical revision of the manuscript for important intellectual content: Amélie Van Eesbeeck, Laura Lannoo, Valérie Race, Philip Van Damme, Kristl G. Claeys, Wim Vandenberghe. Supervision: Wim Vandenberghe, Philip Van Damme, Kristl G. Claeys, Sascha Vermeer.
Acknowledgements: We thank the patients participating in this study. SHVD is funded by a PhD fellowship of the Research Foundation - Flanders (FWO) (1164018N, file number 40900). Research is furthermore funded by the Fund for Clinical Academic studies of the university hospitals of Leuven, Belgium (S61184). KGC and PVD hold the Emil von Behring Chair in Neuromuscular and

Neurodegenerative Disorders by CSL Behring. Several authors of this publication are member of the European Reference Network for Rare Neuromuscular Diseases (ERN Euro-NMD) and of the European Reference Network for Rare Neurological Diseases (ERN-RND).
Conflicts of interest: None.

Chapter 6: *RNF170* mutation causes autosomal dominant sensory ataxia with variable pyramidal involvement

Contributions: Study concept and design: Sien H. Van Daele. Acquisition of data: Sien H. Van Daele. Analysis and interpretation of data: Sien H. Van Daele, Matthieu Moisse, Valérie Race, Amélie Van Eesbeeck, Liesbeth Keldermans, Sascha Vermeer, Hilde Van Esch, Kristl G. Claeys, Philip Van Damme. Drafting the manuscript: Sien H. Van Daele. Critical revision of the manuscript for important intellectual content: Matthieu Moisse, Valérie Race, Amélie Van Eesbeeck, Liesbeth Keldermans, Sascha Vermeer, Hilde Van Esch, Kristl G. Claeys, Philip Van Damme. Supervision: Philip Van Damme.
Acknowledgements: We are grateful to the patients and family members who agreed to participate in this study. SHVD is funded by a PhD fellowship of the Research Foundation - Flanders (FWO) (1164018N, file number 40900). Research is furthermore funded by the Fund for Clinical Academic studies of the university hospitals of Leuven, Belgium (S61184). PVD and HVE are Senior Clinical investigators of FWO. PVD is supported by the ALS Liga België and the KU Leuven funds “Een Hart voor ALS”, “Laeversfonds voor ALS Onderzoek” and the “Valéry Perrier Race against ALS Fund”. KGC and PVD hold the Emil von Behring Chair in Neuromuscular and Neurodegenerative Disorders by CSL Behring.
Conflicts of interest: None.

Curriculum vitae

PERSONAL DATA

BIRTH

01/06/1989, Leuven, Belgium

NATIONALITY

Belgian

EMAIL

sien.vandaele@uzleuven.be

EDUCATION

2022-NOW: MASTER OF SPECIALIST STUDIES IN MEDICINE, HUMAN GENETICS, CATHOLIC UNIVERSITY OF LEUVEN, BELGIUM

2014-2021: MASTER OF SPECIALIST STUDIES IN MEDICINE, NEUROLOGY, CATHOLIC UNIVERSITY OF LEUVEN, BELGIUM

2010-2014: MASTER OF MEDICINE, CATHOLIC UNIVERSITY OF LEUVEN, BELGIUM (MAGNA CUM LAUDE)

2007-2010: BACHELOR OF MEDICINE, CATHOLIC UNIVERSITY OF LEUVEN, BELGIUM (SUMMA CUM LAUDE)

2001-2007: LATIN-MATHEMATICS, MONTFORTCOLLEGE, ROTSELAAR, BELGIUM

EXPERIENCE

2022-NOW: RESIDENT IN HUMAN GENETICS

- 2022-now: University hospital of Leuven, Belgium

2017-NOW: PHD (FWO FUNDING)

- Title: “Genetic determinants of motor neuron vulnerability”
- Laboratory for Neurobiology – Promotor: Prof. Dr. P. Van Damme; Co-promotors: Prof. Dr. K. Claeys, dr. Valérie Race.

2014-2021: RESIDENT IN NEUROLOGY -- GRADUATED AS NEUROLOGIST IN 2021

- 2015-2021: University hospital of Leuven, Belgium
- 2014-2015 : Imelda Hospital, Bonheiden, Belgium

2012: INTERNSHIP IN EXETER, ROYAL DEVON AND EXETER NHS TRUST, UK (SURGERY; 3 MONTHS)

2008: ‘STUDENT-RESEARCHER’ NEUROLOGY (1 MONTH)

- Prof. W. Vanduffel

EXTRA-CURRICULAR

2021	Awarded as ‘Fellow of the European Board of Neurology’ after passing the UEMS-EBN exam with result in the category ‘excellent’ (89%)
2019-2020	Interuniversity certificate in human genetics (BeSHG course)
2019	Statistics course: Statistical analysis of reliability and survival data (KU Leuven)
2017-2018	Statistics course: Univariate data and modelling (KU Leuven)
2017	GCP Refresher course

2014	GCP Interactive Training course
2013-2014	Optional course: Electrocardiography (KU Leuven)
2011-2012	Optional course: Medical English (KU Leuven)
2010-2011	Optional course: Tropical Medicine (KU L)

PUBLICATIONS

Van Daele S.H., Moisse M., Race V. et al. *RNF170* mutation causes autosomal dominant sensory ataxia with variable pyramidal involvement. Eur J Neurol. 2022 Jan;29(1):345-349. doi: 10.1111/ene.15091.

Van Daele S.H., Vermeer S., Van Eesbeeck A. et al. Diagnostic yield of testing for *RFC1* repeat expansions in patients with unexplained adult-onset cerebellar ataxia. J Neurol Neurosurg Psychiatry. 2020 Nov;91(11):1233-1234. doi: 10.1136/jnnp-2020-323998.

Clinical studies

ORARIALS01	Fase 3 study: Arimoclomol in ALS	Sub Investigator
233AS101	Fase 1 study: BIIB67 in ALS-SOD1	Sub Investigator
233AS102	Extension study BIIBO67	Sub Investigator
CY 4031	Fase 3 study: Tirasemtiv in ALS	Sub Investigator

CONFERENCES

International symposium on ALS/MND (Glasgow, 2018)

- poster presentation “The diagnostic yield of NGS in ALS and related neuromuscular disorders”

ADDITIONAL SKILLS

Languages: Dutch (native); English (fluent); French (good).

Computer skills: R statistics (good); SQL (basic-good); Linux (basic); Adobe Illustrator (basic).