New structural variations responsible for Charcot-Marie-Tooth disease: The first two large KIF5A deletions detected by CovCopCan software

Ioanna Pyromali, Alexandre Perani, Angélique Nizou, Nesrine Benslimane, Paco Derouault, Sylvie Bourthoumieu, Mélanie Fradin, Guilhem Sole, Fanny Duval, Constantin Gomes, et al.

Version of Record: https://www.sciencedirect.com/science/article/pii/S2001037021003275
Manuscript_746e320f4b74750d42cecdf1cf2baccc

New Structural Variations Responsible for Charcot-Marie-Tooth Disease: The First Two Large *KIF5A* Deletions Detected by CovCopCan Software

Ioanna Pyromali[1], Alexandre Perani[2], Angélique Nizou[1], Nesrine Benslimane[1], Paco Derouault[3], Sylvie Bourthoumieu[4], Mélanie Fradin[5], Guilhem Sole[6], Fanny Duval[6], Constantin Gomes[7], Frédéric Favreau[1,2], Franck Sturtz[1,2], Corinne Magdelaine[1,2], Anne-Sophie Lia[1,2,3]

[1]Univ. Limoges, MMNP, EA6309, F-87000 Limoges, France

[2]CHU Limoges, Service de Biochimie et de Génétique Moléculaire, F-87000 Limoges, France

[3]CHU Limoges, Service de Bioinformatique, F-87000 Limoges, France

[4]CHU Limoges, Service de Cytogénétique, Génétique Médicale et Biologie de la Reproduction, F-87000 Limoges, France

[5]CHU Rennes, CLAD Ouest, Service de Génétique, F-35203 Rennes, France.

[6]CHU Bordeaux (Groupe Hospitalier Pellegrin), Service de Neurologie et Centre de Référence des Maladies Neuromusculaires AOC, F-33000 Bordeaux, France.

[7]Hôpital Pontchaillou, Département de Neurophysiologie, F-35200 Rennes, France.

Correspondance to: Anne-Sophie Lia,
EA6309 – Neuropathies Périphériques et Maintenance Myélinique,
Faculté de Médecine, Université de Limoges
2 rue du Dr Marcland
87025 Limoges, France
E-mail: anne-sophie.lia@unilim.fr

Abstract

Next-generation sequencing (NGS) allows the detection of mutations in inherited genetic diseases, like the Charcot-Marie-Tooth disease (CMT) which is the most common hereditary peripheral neuropathy. The majority of mutations detected by NGS are single nucleotide variants (SNVs) or small indels, while structural variants (SVs) are often underdiagnosed. *PMP22* was the first gene described as being involved in CMT via a SV of duplication type. To date, more than 90 genes are known to be involved in CMT, with mainly SNVs and short indels described. Herein targeted NGS and the CovCopCan bioinformatic tool were used in two unrelated families, both presenting with typical CMT symptoms with pyramidal involvement. We have discovered two large SVs in *KIF5A*, a gene known to cause axonal forms of CMT (CMT2) in which no SVs have yet been described. In the first family, the patient presented with a large deletion of 12 kb in *KIF5A* from Chr12:57,956,278 to Chr12:57,968,335 including exons 2-15, that could lead to mutation c.(130-943_c.1717-533del), p.(Gly44_Leu572del). In the second family, two cases presented with a large deletion of 3 kb in *KIF5A* from Chr12:57,974,133 to Chr12:57,977,210 including exons 24-28, that could lead to mutation c.(2539-605_*36+211del), p.(Leu847_Ser1032delins33). In addition, bioinformatic sequence analysis revealed that a NAHR (Non-Allelic-Homologous-Recombination) mechanism, such as those in the *PMP22* duplication, could be responsible for one of the *KIF5A* SVs and could potentially be present in a number of other patients. This study reveals that large *KIF5A* deletions can cause CMT2 and highlights the importance of analyzing not only the SNVs but also the SVs during diagnosis of neuropathies.

Keywords: NGS, Structural Variations, CovCopCan, Charcot-Marie-Tooth, *KIF5A*

Abbreviations: ALS = Amyotrophic Lateral Sclerosis; CMT = Charcot-Marie-Tooth; CMT2 = Charcot-Marie-Tooth type 2; CNV = Copy Number Variants; DSMA = Distal-Spinal-Muscular-Atrophy; HSP10 = Hereditary-Spastic-Paraplegia-type-10; NAHR = Non-Allelic Homologous Recombination; NEIMY = Neonatal-Intractable-MYoclonus; NGS = Next Generation Sequencing; SNV = Single Nucleotide Variant; SV = Structural Variant

1. Introduction

Thanks to next-generation sequencing (NGS) technique, we are able to detect pathogenic germline mutations and thus to improve the diagnosis of patients with inherited genetic diseases. Though nowadays, the majority of reported mutations are single nucleotide variants (SNVs), while structural variants (SVs) have rarely been described, probably because the analysis of NGS data is complicated and there are a few available analyzing tools.

Charcot-Marie-Tooth disease (CMT) is the most common hereditary peripheral neuropathy, characterized by damages in both motor and sensory peripheral nerves. *PMP22* duplication was the first mutation to be described in this disease, and is described as explaining 70-80% of the demyelinating form of the disease (CMT1) [1-5]. To date, more than 90 different genes have been reported to be involved in the disease [6]. The majority of the reported mutations in CMT patients consist of point mutations in the coding region of these genes, whereas structural variants (SVs) have only rarely been described [7,8].

KIF5A encodes the heavy chain of Kinesin I (ensembl: ENSG00000155980, UniProtKB-Q12840). This protein is part of a heterotetrameric Kinesin I complex consisting of two light chains and two heavy chains (Fig. 1A). Kinesin I, located in central and peripheral nervous system, is involved in the anterograde transport of various cargoes (RNA, mitochondria, neurofilaments) along microtubules in neurons [9]. The kinesin heavy chain subunit, encoded by *KIF5A*, presents three main domains: motor (amino acid residue 9-327), stalk (amino acid residue 328-906), and tail (amino acid residue 907-1032) (Fig. 1B). The motor domain possesses ATPase activity necessary to produce kinetic force, the stalk domain allows the dimerization with another kinesin heavy chain forming a coiled-coil interaction, and the tail interacts with the kinesin light chain and cargoes [10-12].

Missense mutations located on the Kinesin-I heavy chain motor domain were first described as mainly causing a rare form of Hereditary-Spastic-Paraplegia type 10 (HSP10; OMIM#604187) [13] but also Charcot-Marie-Tooth type 2 disease (CMT2; ORPHA:324611) [14]. The clinical symptoms of HSP observed in patients presenting with *KIF5A* mutations are pyramidal syndrome and sphincter disturbances, while CMT signs consist of lower and/or upper limb weakness with gait disturbances, sensory deficits, depressed tendon reflexes, articular deformations (*pes cavus* and scoliosis) and distal muscular atrophy [15,16]. In addition, other *KIF5A* mutations have also been identified in a severe neonatal neurological syndrome called NEIMY (Neonatal-Intractable-MYoclonus; OMIM#617235) [17], in Amyotrophic-Lateral-Sclerosis (ALS, OMIM#617921) [18,19], and recently in Distal Spinal Muscular Atrophy (DSMA) (Fig. 1C) [20]. Thus, *KIF5A* mutations present a wide clinical spectrum exhibited in several diseases associated with an autosomal dominant mode of inheritance. Interestingly, to date, only SNVs or frameshift mutations have been described in *KIF5A*. To our knowledge, structural variations have never been discovered.

Herein, we describe the first two structural variants in *KIF5A* responsible for CMT symptoms associated with pyramidal syndrome in two distinct families. Targeted NGS and CovCopCan bioinformatic tool [21] revealed a 12 kb deletion in *KIF5A* from Chr12:57,956,278 to Chr12:57,968,335 including exons 2-15 in the first family and a 3 kb deletion in *KIF5A* from Chr12:57,974,133 to Chr12:57,977,210 including exons 24-28 in the second family. These SVs have not previously been described and highlight the importance of SV analysis of NGS data in the diagnosis of neuropathies.

2. Materials and methods

2.1. Patients

Family 1: The proband (A), a 60-year-old man presented axonal CMT syndrome associated with several HSP10 signs (Fig. 2A-Family 1). Family 2: The proband (B), a 70-year-old man, and his 47-year-old son (C) presented axonal CMT syndrome associated with atypical signs. The proband B's spouse was healthy (Fig. 2B-Family 2).

Peripheral blood was collected into EDTA tubes for each patient after informed consent was obtained according to the Declaration of Helsinki. DNA extraction was then performed using standard methods (Illustra-DNA-Extraction-kit-BACC3, GEHC). The study has been approved by the ethics committee of Limoges University Hospital (n° 387-2020-43).

2.2. Next Generation Sequencing (NGS)

NGS strategy was performed using a 93-gene custom panel designed for diagnosis of CMT and associated neuropathies (as described by Miressi *et al.* [22]). The amplified library was prepared with Ion-P1-HiQ-Template-OT2-200 kit (Ampliseq-Custom; Life Technologies), sequenced on Ion-Proton sequencer (Life-Technologies), and mapped to the human-reference-genome GHCh37.

2.3. Bioinformatics analysis

Variants were evaluated with Alamut-Visual-Interpretation Software v.2.11 (Interactive-Biosoftware, Rouen, France) using the NM_004984.4 reference sequence for the *KIF5A* gene. Databases such as gnomAD (https://gnomad.broadinstitute.org/), dbSNP135 (National-Center-Biotechnology-Information [NCBI],

http://www.ncbi.nlm.nih.gov/projects/SNP/) and Clin-Var (www.ncbi.nlm.nih.gov/clinvar) were used as well.

Transposable elements were detected from the genomic sequence of the *KIF5A* using RepeatMasker software (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker) with default settings. The graph was then generated with the Gviz package [23].

2.4. Structural Variation detection

SVs were detected using Cov'Cop and CovCopCan analysis tools starting from the coverage file provided after NGS sequencing by Ion-Proton-sequencer [21, 24]. Cov'Cop and CovCopCan software programs use a two-stage correction and normalization algorithm to identify unbalanced SVs, such as CNVs (Copy Number Variants), using NGS read depth. These software are based on the principle that in normal cases, both alleles should be amplified similarly within each amplicon. To the normal cases a theoretical score of 1 is attributed, whereas deletions or duplications are revealed by low (<0.5) or high (>1.5) scores respectively. At least three adjacent amplicons with altered values were required in order to highlight a CNV. CovCopCan software also allows graphical representation of data.

In order to delineate the exact deletion breakpoints, long range PCRs were performed using Taq-Polymerase-Phusion-hot-Start II (Qiagen), according to the manufacturer's recommendations. Several long range PCRs were tested using forward primers in exon 1 or all along intron 1 and reverse primer in exon 16 for Family 1 in order to optimize the detection of the deleted allele. Only one PCR was tested for family 2. The final following primers (Sigma Aldrich) were used: Int1-F (AGTCAGTCACCTATGATAGAG) and Ex16-R (GCCACAGTGAACTCCTCC) for Family 1; Int22-F (CTGCCCCTGTTGCCCCTATG) and Ex29-R (GCCGCTGGAGAATCTTGGTCT) for Family 2. Conditions of PCR were 98°C 1

min, followed by 30 cycles [98°C 1 min; 59°C 30 sec (Fam 1) and 64°C 30 sec (Fam 2); 72°C 8 min (Fam 1) and 72°C 4 min (Fam 2)] (See Supplementary data 1 for details on PCR results). PCR products were sequenced by the Sanger sequencing method using the Big Dye Terminator Cycle Sequencing Kit v2 (ABI Prism, Applied Biosystems), but also a walking primer strategy for Family 2 (See Supplementary data 1 for details on primers used to identified the breakpoints)

3. Results

3.1. Atypical clinical phenotype

Regarding Family-1 (Fig. 2A), patient A was a 60-year-old man, presenting gait disturbances and distal muscle wasting without objective motor weakness of the four limbs since the age of 26. He had *pes cavus* and pyramidal syndrome. This patient has also been evaluated using (i) the Charcot-Marie-Tooth Neuropathy Score (CMTN) with a value at 2/4, (ii) the Medical Research Council (MRC) with a score 100/100 and (iii) Overall Neuropathy Limitation Scale (ONLS) with a score 3/12. Electroneuromyography showed a sensorimotor axonal peripheral neuropathy with fasciculations on all four limbs (Fig. 2C). His large biological analysis remains normal, with a normal B12 level of 432 pmol/L (normal range:187-885) and no diabetes (HbA1C : 5.2%, with normal value expected around 6%). Patient A's father deceased at 80 years old without showing any neurological symptoms, while his mother deceased at the age of 75, having presented cognitive disorders and a ventricular cyst. Neither of the proband's five siblings, nor his two sons (38 and 30 years old), presented with any neurological problems.

Regarding Family-2 (Fig. 2B), patient B was a 70-year-old man, presenting with gait disturbances with *musculum tibialis anterior* weakness for at least ten years, fingertip paraesthesia and mild tremor. Articular deformities with *pes cavus* and mild scoliosis were also noted. He can however still do 12 km long walks. Axonal CMT disease with sensorimotor axonal peripheral neuropathy (low amplitude, subnormal velocity) was suspected based on electroneuromyography (Fig. 2C). Furthermore, a pyramidal syndrome with brisk deep tendon reflexes in the lower limbs was reported, associated with sphincter disturbance. He had no siblings, and his parents were deceased without neurological symptoms. He had two sons, of whom, one (C) also presented CMT symptoms.

Patient C was a 47-year-old man, presenting an axonal CMT disease with gait disturbance since the age of 40. He complains of difficulties walking since several years but only describes one fall. He had *pes cavus,* and electroneuromyography showed a sensorimotor axonal peripheral neuropathy (Fig. 2C). This symptomatology was also associated with brisk deep tendon reflexes in the lower limbs and moderate tremor. Proband B's spouse was healthy, as was his second son. However, his DNA was not available.

None of the patients presented other sensory deficits (neither vision loss nor deafness) and they do not have medical history of medication/intoxication that could be related to their CMT symptoms.

3.2. Detection of Structural Variants using CovCopCan

Targeted NGS analysis of our 93-gene panel involved in peripheral neuropathies was performed on DNA from patients A and B DNA. CNVs detection was performed using the user-friendly CovCopCan tool and allow the discovery of two heterozygous deletions for both cases. For patient A, CovCopCan analysis revealed a deletion of 17 adjacent amplicons covering the genomic region Chr12:57,957,118-Chr12:57,966,569 including *KIF5A* exons 2-15 (Fig. 3A). Regarding patient B, a deletion of 7 adjacent amplicons was identified, covering the genomic region Chr12:57,974,529-Chr12:57,977,002 including *KIF5A* exons 24-28 (Fig. 3B). In comparison, no variation has been observed in control patients (Fig. 3C and 3D), in which normal values were observed for *KIF5A*. No additional positive variant candidate to explain patients' phenotype (SNVs and short indels) has been identified for patients using the standard NGS analysis.

3.3. Configuration of Structural Variants

Long range PCR, from intron 1 to exon 16, was then performed on DNA from patient A (see supplementary data 1). We amplified and sequenced the deleted allele to identify the exact breakpoints positions at Chr12:57,956,278 and Chr12:57,968,335 (Fig. 3E), corresponding to a 12,057 base pairs deletion. This *KIF5A* deletion identified on patient A DNA starts in intron 1 at position c.130-943 and ends in intron 15 at position c.1717-533. On KIF5A protein, it would correspond to the whole kinesin I heavy chain motor domain and part of the stalk domain and could be named p.(Gly44_Leu572del) (Fig. 3G, Family-1).

The same strategy performed, from intron 22 to exon 29, on patient B DNA allowed to amplify (see supplementary data 1) and sequence the deleted allele. We identified the breakpoints positions at Chr12:57,974,133 and Chr12:57,977,210 (Fig. 3F), corresponding to a 3,077 base pairs deletion. The identical result was obtained for the son (C). This *KIF5A* deletion identified on patients B and C DNAs starts in intron 23 at position c.2539-605 and ends in exon 29 at position c.*36+211. On KIF5A protein, it would affect the end of the Kinesin-I heavy chain stalk domain and tail domain and potentially alter its activity and could be named p.(Leu847_Ser1032delins33) (Fig. 3G, Family-2).

3.4. Informatics analysis of sequences and detection of the recombination mechanism

Then we proceed to the detection of the transposable elements presented in the genomic sequence of the *KIF5A*, using RepeatMasker software. This analysis revealed that the breakpoints of the exons 2-15 deletion (patient A) are located in a region rich in Alu repeat elements (Fig. 4). The numerous Alu repeat sequences (AluSx and AluJb) in this region could promote an erroneous recombination and thus a large gene deletion due to a non-allelic homologous recombination (NAHR) event.

4. Discussion

In this paper, we highlight the importance of bioinformatic analysis of NGS data in order to detect new pathogenic structural variants. For the first time, we report here that large *KIF5A* deletions, detected in two different families, may be responsible for CMT atypical cases associated with pyramidal syndrome. These deletions are carried in an autosomal dominant manner. In the first family, the patient presented with a large deletion of 12 kb in *KIF5A* from Chr12:57,956,278 to Chr12:57,968,335, that could lead to mutation c.(130-943_c.1717-533del), including the deletion of exons 2 to15. In the second family, two cases presented with a large deletion of 3 kb in *KIF5A* from Chr12:57,974,133 to Chr12:57,977,210, that could lead to mutation c.(2539-605_*36+211del), including the deletion of exons 24 to 28. Both of the deletions were detected by using the CovCopCan analysis tool.

KIF5A large deletions have not been previously described. Indeed, to date, according to the Human-Gene-Mutation-Database (HGMD®) Professional 2020.3 (Qiagen), the published mutations mainly consist of SNVs leading to missense mutations, splicing alterations, premature stop codon and short deletions that may result in frame-shift. We showed here that *KIF5A* large deletions could be involved in atypical CMT cases.

It is important to remind that *KIF5A* encodes the heavy chain of kinesin I, which interacts with another heavy chain in order to form a homodimer. This homodimer interacts then with two light chains of kinesin I in order to form an active heterotetrametric complex. If one of this heavy chain is mutated but is still able to interact with another monomer including the wild type one (WT), only 25% of the dimer would be WT-WT and completely active. If the mutation has an effect on the motor domain or the cargoes attachment, this means that 75% of the remaining dimers (WT-Mutated and Mutated-Mutated) would have a disturbed activity. This kind of dominance mechanism has already been shown on other dimeric proteins such as TNSALP involved in hypophosphatasia [25, 26].

In the previous studies, it has been shown that the clinical phenotype seems to be dependent on the mutated KIF5A domain as shown in figure 1, with severe phenotype observed when located in the tail domain and milder ones when located in motor or stalk domains. It is also interesting to note that indirectly some SNVs, previously described, such as splicing alterations, short deletions and premature stop codons could lead to a truncated protein, that could be compared to the deletions we described here. Interestingly, splicing mutations predicted or shown to result on exon deletion would lead to mild symptoms when located in the area coding the motor domain [27, 28] and to a severe one when located in the area coding the tail domain ([18, 19], for instance). In addition, small deletions leading to a frameshift have only been described, to our knowledge, in the same area (Exons 24 to 27) coding the tail domain (HGMD data) and are always associated to severe phenotype such as NEIMY and ALS. These observations could suggest that, as for point mutations (Fig. 1), functional domain deleted in KIF5A dictates the clinical phenotype.

In contrast, the rare described non-sense mutations, located in exons 20 or 24 [29, 30], lead to CMT or HSP symptoms, as those observed in our patients. In these non-sense mutations cases, if the protein KIF5A is produced, the tail part would be missing and one could expect severe symptoms, while milder symptoms are observed. In these cases, mutated RNA are probably degraded by non-sense-mediated-decay mechanism (NMD), leading to the absence of the truncated heavy chain of Kinesin-I. The cell would then produce only normal KIF5A, that would create functional homodimer, but in a smaller proportion (50%). This could lead to a mild phenotype by a possible haplo-insufficiency effect.

In the cases, we described here, the deletions could result in (i) the production of a truncated protein or to (ii) in the complete absence of the heavy chain of Kinesin-I. (i) Regarding the first hypothesis, Alamut-Visual-Interpretation-Software was used in order to check for any reading frame alteration resulting from exons' deletion for both families. This

analysis revealed that, for patient A (c.(130-943_1717-533del) including exons 2-15 deletion), assuming correct splicing occurs between exon 1 and 16, a correct reading frame could be preserved, but would produce a mutated heavy chain p.(Gly44_Leu572del) lacking almost all of its motor domain and part of its stalk domain (Fig. 3E). For patients B and C, the deletion c.(2539-605_*36+211del), including to exons 24 to 28, removes the last coding exon (exon 28) but not the non-coding exon 29. Again, if the splicing is performed correctly between exon 23 and 29, the translation of exon 29 would lead to 33 additional amino acids on a mutated heavy chain p.(Leu847_Ser1032delins33) in which the end of the stalk domain and the entire classic tail would be missing (Fig. 3E). For the first deletion, if the truncated heavy chain exists, it would probably not interact with the wild-type one due to the large loss of the stalk domain. Wild-type functional homodimer would be then present in smaller proportion (50%). This could lead to a mild phenotype comparable to those observed in patients harboring a non-sense mutation in which haplo-insufficiency could be suggested. For the second deletion, if the truncated heavy chain exists, it could have either a dominant negative effect by interacting with the wild type heavy chain and thus disrupting kinesin function, or an effect on itself (gain of function effect) due to the possible 33 additional amino-acids. This would be comparable to splicing mutations and the small deletions leading to a frameshift in the tail of KIF5A and to a severe phenotype (ALS, NEIMY). However, in our case (Family 2), the symptoms are milder suggesting a different pathophysiology mechanism. (ii) A more likely second hypothesis would be that the kinesin heavy chain derived from the deleted alleles would be absent (not produced or rapidly degraded). In this case, the deletions would also have a haplo-insufficiency effect, and would be classified as PVS1 according to the ACMG guidelines [31], since it corresponds to multi-exons deletions in a gene where loss of function can be a mechanism of the disease. Even if these deletions seem to be the cause of the patients' disease, their underlying pathophysiologies are difficult

to decipher. Indeed, *KIF5A* being expressed mainly in the nervous system, it is difficult reach a conclusion due to the lack of nervous tissue availability.

In addition, regarding the appearance of these large deletions, we noticed that the parents of both patients A and B were deceased elderly without known neurological symptoms. Patients A and B thus seem to be the first members of their families presenting symptoms due to *KIF5A* partial deletion. We therefore hypothesize that these CNVs probably appeared *de novo*. Analysis of the exact breakpoints position of the deletion 2 (exons 24-28) did not allow us to detect any particular joint mechanism. Interestingly, for the first deletion (exons 2-15), a mechanism similar to that involved in *PMP22* duplication could be suspected. This kind of genomic alteration, due to a NAHR event, has already been described in several hereditary diseases including CMT disease, particularly in *PMP22* duplication [32]. In patient A, it seems that erroneous recombination occurred within the AluSx sequences (13 identical bp) via a NAHR mechanism. Because of the numerous repetitive elements present in this *KIF5A* region (Fig. 4), we believe that partial deletions in *KIF5A* could be currently underestimated, and we suggest searching for them systematically when performing NGS sequencing on CMT-patients who also harbor a pyramidal syndrome.

5. Conclusion

In conclusion, our approach using NGS data analysis to look for small nucleotide variants (SNVs or short indels) combined with investigation for large CNVs (using the CovCopCan tool for example) allowed us to highlight the existence of two different large structural variants in two distinct families. The clinical signs in these patients are in agreement with a variation of *KIF5A* and they did not present any other mutations in genes currently associated with CMT, leading to the conclusion that those structural variants are indeed the cause of their disease. In our cohort of 765 CMT patients, we identified six patients with *KIF5A* variations, of whom two presented these SVs on *KIF5A*, *i.e.* in 33%, suggesting that this kind of variation is present in a significant manner. We conclude so that also large deletions in *KIF5A* are associated to CMT/HSP10 diseases.

We believe our strategy of looking systematically for SVs when performing NGS analysis could increase the diagnosis rate of patients suffering from neuropathies. This strategy could also be employed by molecular biologists and geneticists to improve all inherited diseases diagnosis.

6. Author Statement

All authors have reviewed and edited the manuscript

IP: participated to the bioinformatics study, performed experiments and wrote most of the manuscript

AP: participated to the bioinformatics study, performed experiments and proofread the manuscript

AN: performed experiments and proofread the manuscript

NB: performed experiments and proofread the manuscript

PD: participated to the bioinformatics study and proofread the manuscript

SB: performed experiments

MF: reported clinical study

GS: reported clinical study

FD: reported clinical study

CG: reported clinical study

FF: performed critical revision of the manuscript

FS: performed critical revision of the manuscript

CM: planned the experiments, contributed to the interpretation of the results and proofread the manuscript

ASL: initiated the work, designed and planned the experiments, analyzed the results and proofread the manuscript

Declaration of Competing Interest

The authors declare no competing financial interest or personal relationships that could have influenced the content of this article.

Acknowledgement

We would like to thank the technicians in the Biochemistry and Molecular Genetics Laboratory and the Sequencing and Bioinformatic platform at Limoges CHU, especially Emilie Guerin, Valentin Tilloy and Sophie Alain for their excellent technical support.

References :

1. Lupski JR, de Oca-Luna RM, Slaugenhaupt S, Pentao L, Gizzetta V, *et al.* DNA Duplication Associated with Charcot-Marie-Tooth Disease Type 1A. Cell 1991; 66 (2):219-32. doi: 10.1016/0092-8674(91)90613-4.
2. Matsunami N, Smith B, Ballard L, Lensch MW, Robertson M, *et al*. Peripheral Myelin Protein-22 Gene Maps in the Duplication in Chromosome 17p11.2 Associated with Charcot-Marie-Tooth 1A. Nat Genet. 1992; 1 (3):176-9. doi: 10.1038/ng0692-176.
3. Valentijn LJ, Bolhuis PA, Zorn I, Hoogendijk JE, van den Bosch N, *et al*. The Peripheral Myelin Gene PMP-22/GAS-3 is Duplicated in Charcot-Marie-Tooth Disease Type 1A. Nat Genet. 1992; 1(3):166-70. doi: 10.1038/ng0692-166.
4. Timmerman V, Nelis E, Van Hul W, Nieuwenhuijen BW, Chen KL, *et al*. The Peripheral Myelin Protein Gene PMP–22 Is Contained within the Charcot–Marie–Tooth Disease Type 1A Duplication. Nat Genet 1992; 1 (3): 171–75. doi: 10.1038/ng0692-171
5. Murphy SM., Laura M, Fawcett K, Pandraud A, Liu YT, *et al*. Charcot-Marie-Tooth Disease: Frequency of Genetic Subtypes and Guidelines for Genetic Testing. J Neurol Neurosurg Psychiatry 2012; 83 (7): 706–10. doi: 10.1136/jnnp-2012-302451
6. Stojkovic, T. Hereditary Neuropathies: An Update. [Review]. Rev Neurol 2016; 172 (12): 775–78. doi: 10.1016/j.neurol.2016.06.007
7. Gonzaga-Jauregui C, Zhang F, Towne CF, Batish SD, Lupski JR. GJB1/Connexin 32 Whole Gene Deletions in Patients with X-Linked Charcot–Marie–Tooth Disease. Neurogenetics 2010; 11 (4): 465–70. doi: 10.1007/s10048-010-0247-4
8. Mortreux J, Bacquet J, Boyer A, Alazar E, Bellance R, *et al*. Identification of Novel Pathogenic Copy Number Variations in Charcot-Marie-Tooth Disease. J Hum Genet 2020; 65 (3): 313–23. doi: 10.1038/s10038-019-0710-5
9. Niclas J, Navone F, Hom-Booher N, Vale RD. Cloning and localization of a conventional kinesin motor expressed exclusively in neurons. Neuron 1994; 12: 1059-1072. doi: 10.1016/0896-6273(94)90314-x
10. Vale RD, Reese TS, Sheetz MP. Identification of a Novel Force-Generating Protein, Kinesin, Involved in Microtubule-Based Motility. Cell 1985; 42 (1): 39–50. doi: 10.1016/s0092-8674(85)80099-4
11. Kull FJ, Sablin EP, Lau R, Fletterick RJ, Vale RD. Crystal Structure of the Kinesin Motor Domain Reveals a Structural Similarity to Myosin. Nature 1996; 380 (6574): 550–55. doi: 10.1038/380550a0
12. Hirokawa N, Noda Y, Tanaka Y, Niwa S. Kinesin Superfamily Motor Proteins and Intracellular Transport. [Review]. Nat Rev Mol Cell Biol; 10 (10): 682–96. doi: 10.1038/nrm2774
13. Reid E, Kloos M, Ashley-Koch A, Hughes L, Bevan S, *et al*. A kinesin heavy chain (KIF5A) mutation in hereditary spastic paraplegia (SPG10). Am J Hum Genet 2002; 71: 1189–94. doi: 10.1086/344210.
14. Liu YT, Laurá M, Hersheson J, Horga A, Jaunmuktane Z, *et al*. Extended Phenotypic Spectrum of KIF5A Mutations. Neurology 2014; 83 (7): 612–19. doi: 10.1212/WNL.0000000000000691
15. Fink JK. Hereditary spastic paraplegia. Curr Neurol Neurosci Rep 2006: 6 (1): 65–76. doi: 0.1007/s11910-996-0011-1

16. Bird TD. Charcot-Marie-Tooth (CMT) Hereditary Neuropathy Overview. 1998 Sep 28 [Updated 2021 Mar 18]. In: Adam MP, Ardinger HH, Pagon RA, et al., editors. GeneReviews® [Internet].
17. Duis J, Dean S, Applegate C, Harper A, Xiao R, *et al*. KIF5A Mutations Cause an Infantile Onset Phenotype Including Severe Myoclonus with Evidence of Mitochondrial Dysfunction. Ann Neurol 2016; 80 (4): 633–37. doi: 10.1002/ana.24744
18. Nicolas A, Kenna KP, Renton AE, Ticozzi N, Faghri F, *et al*. Genome-Wide Analyses Identify KIF5A as a Novel ALS Gene. Neuron 2018; 97 (6): 1268-1283.e6. doi: 10.1016/j.neuron.2018.02.027.
19. Brenner D, Yilmaz R, Müller K, Grehl T, Petri S, *et al*. Hot-Spot KIF5A Mutations Cause Familial ALS. Brain 2018; 141 (3): 688–97. doi: 10.1093/brain/awx370.
20. de Fuenmayor-Fernández de la Hoz CP, Hernández-Laín A, Olivé M, Sánchez-Calvín MT, Gonzalo-Martínez JF, *et al*. Adult-onset distal spinal muscular atrophy: a new phenotype associated with KIF5A mutations. Brain 2019; 142(12):e66. doi: 10.1093/brain/awz317
21. Derouault P, Chauzeix J, Rizzo D, Miressi F, Magdelaine C, *et al*. CovCopCan: An Efficient Tool to Detect Copy Number Variation from Amplicon Sequencing Data in Inherited Diseases and Cancer. PLoS Comput Biol 2020; 16 (2): e1007503. doi: 10.1371/journal.pcbi.1007503
22. Miressi F, Faye PA, Pyromali I, Bourthoumieu S, Derouault P, *et al*. A mutation can hide another one: Think Structural Variants! Comput Struct Biotechnol J 2020; 18:2095-2099. doi: 10.1016/j.csbj.2021.01.037
23. Hahne F, Ivanek R. Statistical Genomics: Methods and Protocols. In Mathé E, Davis S (eds.), chapter Visualizing Genomic Data Using Gviz and Bioconductor, 335–351. Springer New York, New York, NY. 2016; ISBN 978-1-4939-3578-9, doi: 10.1007/978-1-4939-3578-9_16,
24. Derouault P, Parfait B, Moulinas R, Barrot CC, Sturtz F, *et al*. 'COV'COP' Allows to Detect CNVs Responsible for Inherited Diseases among Amplicons Sequencing Data. Bioinformatics (Oxford, England) 2017; 33 (10): 1586–88. doi: 10.1093/bioinformatics/btx017
25. Lia-Baldini AS, Muller F, Taillandier A, Gibrat JF, Mouchard M, *et al*.. A molecular approach to dominance in hypophosphatasia. Human Genet 2001; 109(1):99-108. doi: 10.1007/s004390100546.
26. Lia-Baldini AS, Brun-Heath I, Carrion C, Simoon-Bouy B, Serre JL, *et al*. A new mechanism of dominance in hypophosphatasia: the mutated protein can disturb the cell localization of the wild-type protein. Human Genet 2008; 123(4):429-32. doi: 10.1007/s00439-008-0480-1.
27. Schüle R, Kremer BP, Kassubek J, Auer-Grumbach M, Kostic V, *et al*. SPG10 is a rare cause of spastic paraplegia in European families. J Neurol Neurosurg Psychiatry. 2008; 79(5):584-7. doi: 10.1136/jnnp.2007.137596.
28. van de Warrenburg BP, Schouten MI, de Bot ST, Vermeer S, Meijer R, *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. doi: 10.1038/ejhg.2016.42
29. Hsu YH, Lin KP, Guo YC, Tsai YS, Liao YC, *et al*. Mutation spectrum of Charcot-Marie-Tooth disease among the Han Chinese in Taiwan. Ann Clin Transl Neurol. 2019; 6(6):1090-1101. doi: 10.1002/acn3.50797.

30. Lynch DS, Koutsis G, Tucci A, Panas M, Baklou M, *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. doi: 10.1038/ejhg.2015.200.
31. Richards S, 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. Genet Med. 2015 May ; 17(5): 405–424.
32. Lupski JR. Charcot-Marie-Tooth disease: Lessons in genetic mechanisms. [Review]. Mol Med 1998; 4:3–11.

Legends

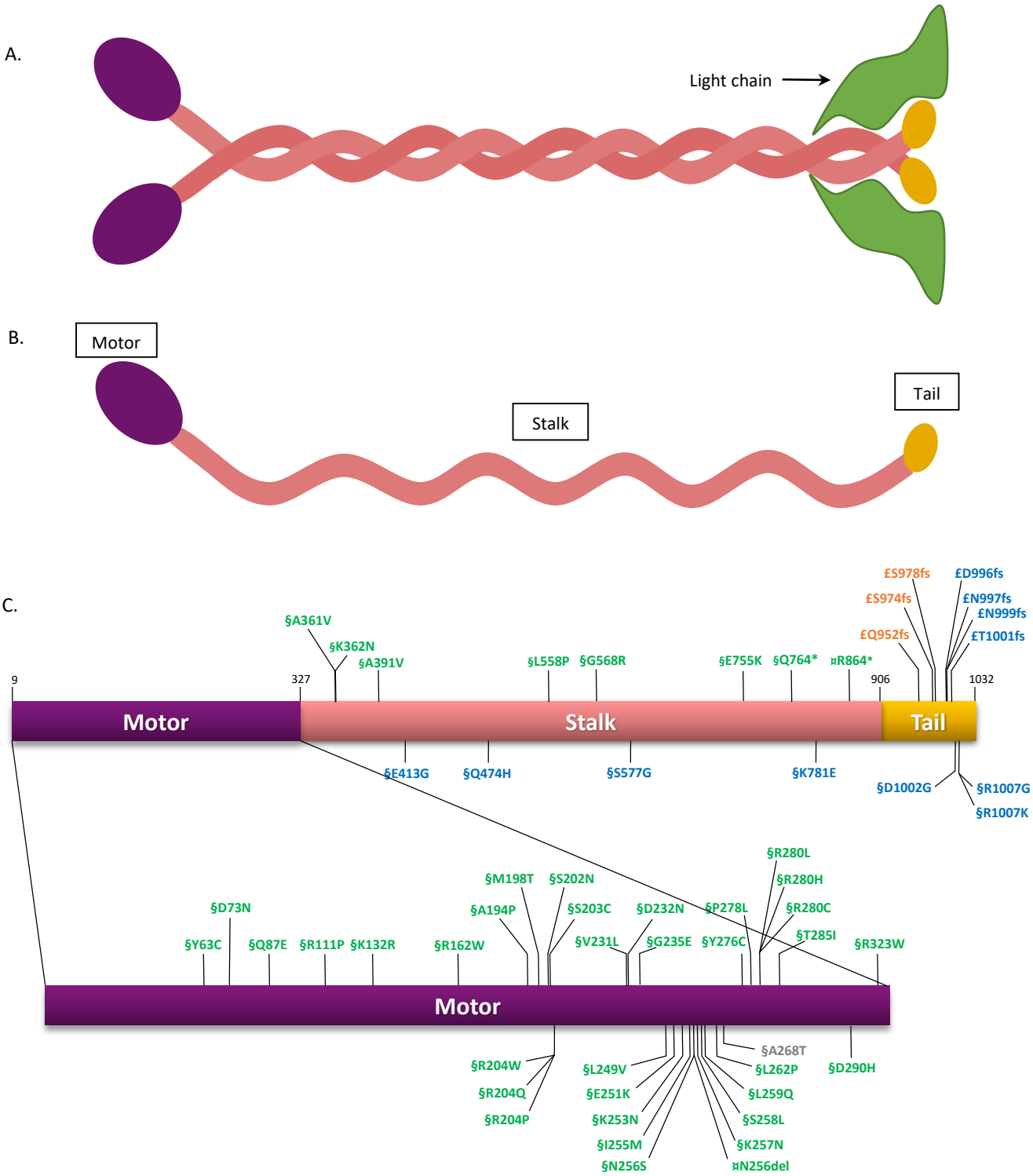
Fig. 1. A) Kinesin-I structure. Two heavy chains and two light chains (in green) form the heterotetramer of Kinesin-I. B) Kinesin-I heavy chain encoded by *KIF5A*. The heavy chain of Kinesin-I contains three domains: a motor domain (purple), a stalk domain (pink) and a tail domain (yellow). C) Location of known mutations in *KIF5A*. Functional domains are represented. Each mutation is indicated by the number of the affected amino-acid. CMT2 isolated or associated with HSP and isolated HSP are shown in green, ALS are shown in blue, NEIMY are shown in orange and DSMA are shown in grey. § = missense mutation; ¤ = deletion; £ = frame-shift mutations. 3'-UTR mutations, and splice mutations affecting the tail are not shown.

Fig. 2. Family pedigrees with segregation of the phenotypes and genotypes and neurophysiological recordings. A) Pedigree of family 1 and B) Pedigree of family 2. (d.) indicates the age of death, (+) indicates normal *KIF5A* allele and (Del1 or Del2) allele with a *KIF5A* partial deletion; Del1 corresponding to c.(130-943_1717-533del) including exons 2-15 and Del2 corresponding to c.(2539-605_*36+211del) including exons 24-28. C) Patients

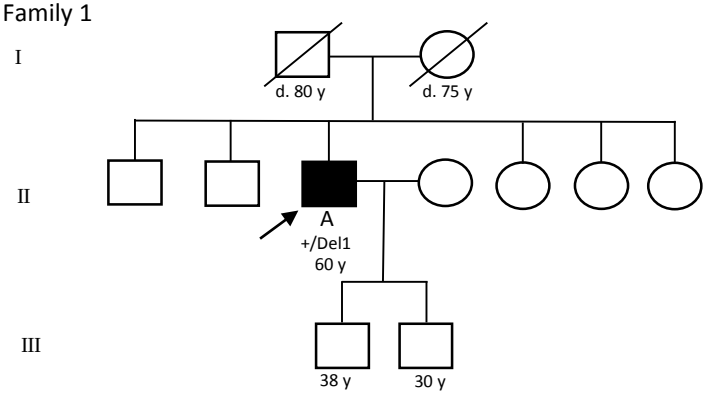
A, B and C neurophysiological recordings. Abnormal values are represented in bold (Vel: Velocity; Amp: Amplitude; NR: no response; ND: not determined).

Fig. 3. CovCopCan analysis of *KIF5A* and surrounding genes on chromosome 12 and Sanger Sequencing results. A, B), C) and D) CovCopCan graphical representation of patients A and B analyses and controls samples respectively. Each dot corresponds to an amplicon, which are distributed along the x-axis in accordance with their genomic position. The y-axis corresponds to the normalized values of each amplicon. Grey dots correspond to values considered "normal" (value around 1), while red or orange dots show the amplicons' duplication (value around 1.5) or deletion (value around 0.5) respectively. The deletion areas are highlighted with yellow rectangles. E) *KIF5A* intron 1 and intron 15 Sanger sequencing results. Results reveal the break points at positions Chr12:57,956,278 and Chr12:57,968,335, corresponding to a 12,057 base pairs deletion. F) *KIF5A* intron 23 and intron 28 Sanger sequencing results. Results reveal the break points at positions Chr12:57,974,133 and Chr12:57,977,210, corresponding to a 3,077 base pairs deletion. (G) Localization of both deletions on the protein encoded by *KIF5A*.

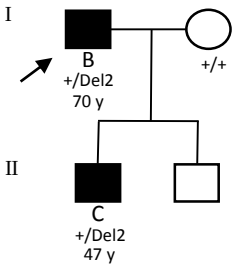
Fig. 4. Mapping of *KIF5A* structural variants. Orange rectangles represent *KIF5A* exons. Red, green and blue arrows represent transposable elements of Alu family: AluY, AluJb and AluSx respectively. Breakpoints of the deletion of Family 1 are situated in a region containing a large number of repetitive Alu sequences, promoting a mistaken recombination event via the NAHR mechanism.


A.
Light chain
B.
Motor
Stalk
Tail
C.
£S978fs
£D996fs
£S974fs
£N997fs
£N999fs
£Q952fs
£T1001fs
§A361V
§K362N
§A391V
§L558P
§G568R
§E755K
§Q764*
¤R864*
9
327
906
1032
Motor
Stalk
Tail
§E413G
§Q474H
§S577G
§K781E
§D1002G
§R1007G
§R1007K
§R280L
§M198T
§S202N
§R280H
§D73N
§A194P
§S203C
§D232N
§P278L
§R280C
§Y63C
§Q87E
§R111P
§K132R
§R162W
§V231L
§G235E
§Y276C
§T285I
§R323W
Motor
§A268T
§R204W
§L249V
§L262P
§D290H
§R204Q
§E251K
§L259Q
§R204P
§K253N
§S258L
§I255M
§K257N
§N256S
¤N256del

A. Family 1



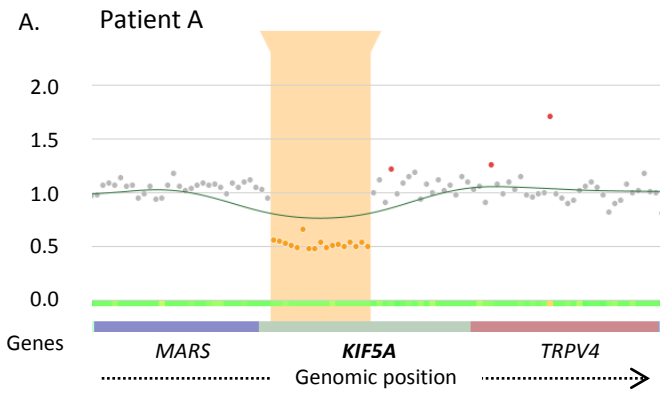
B. Family 2

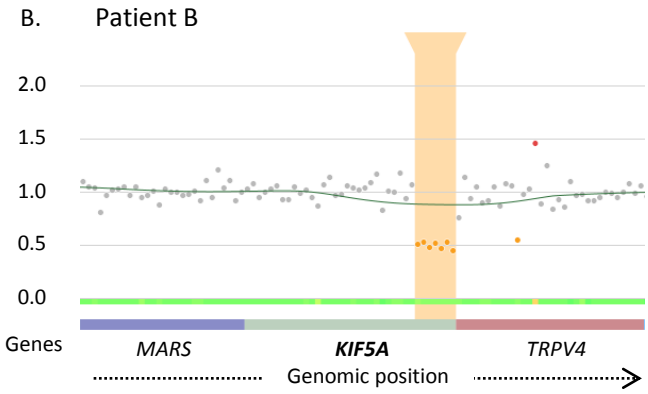


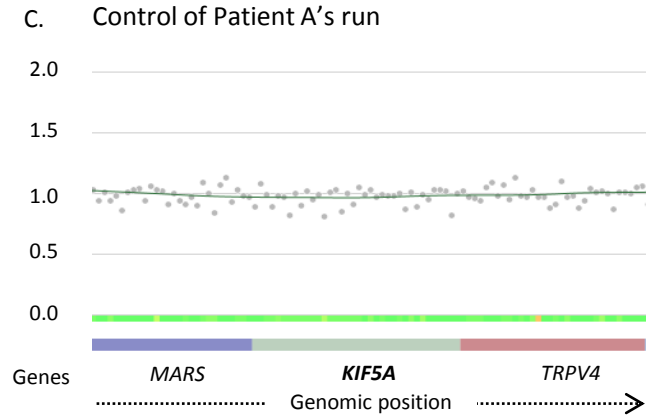
C.

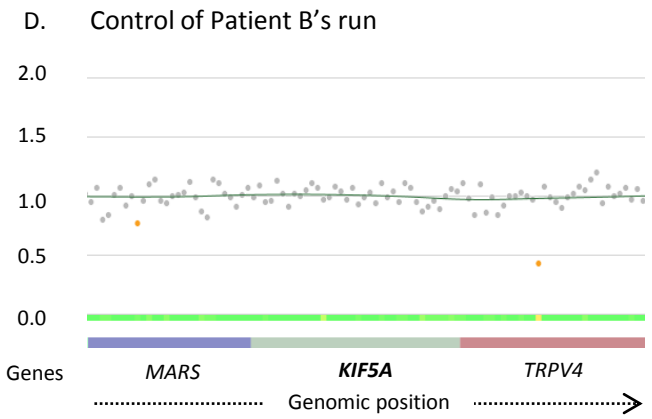
	Motor Nerve Conduction Values (MNCV)						Sensory Nerve Conduction Values (SNCV)					
	Median		Fibular		Tibial		Median		Radial		Sural	
	Vel (m/s)	Amp (mV)	Vel (m/s)	Amp (mV)	Vel (m/s)	Amp (mV)	Vel (m/s)	Amp (μV)	Vel (m/s)	Amp (μV)	Vel (m/s)	Amp (μV)
Patient A	54.6	10.1	**33.3**	**1.1**	42.3	3.8	**44.6**	**17.0**	51.4	20.0	**36.2**	**4.2**
Patient B	51.6	9.7	**NR**	**NR**	**36.8**	**2.8**	**39.2**	**8.5**	**44**	**8.0**	**29.2**	**4.3**
Patient C	54.6	4.5	**40.2**	2.3	**40.7**	5.9	ND	ND	ND	ND	**35.7**	**1.8**

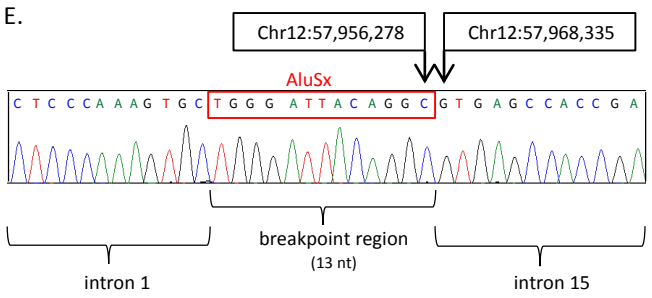
Normal values of MNCV: median (>49m/s; >4mV), fibular (>41m/s; >2mV), tibial (>41m/s; >3mV)
Normal values of SNCV: median (>53m/s; >20uV), radial (>48m/s; >20uV), sural (>41m/s; >6uV)

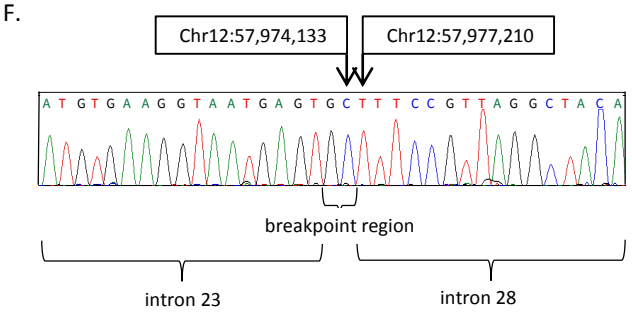

A.
Patient A
2.0
1.5
1.0
0.5
0.0
Genes
MARS
KIF5A
TRPV4
Genomic position

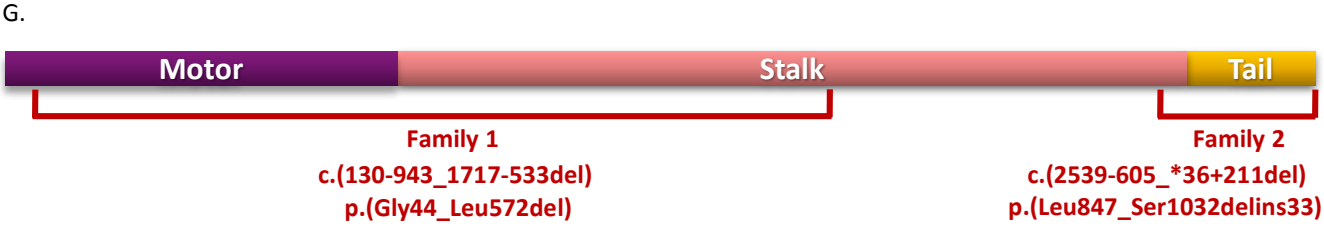

B.
Patient B
2.0
1.5
1.0
0.5
0.0
Genes
MARS
KIF5A
TRPV4
Genomic position

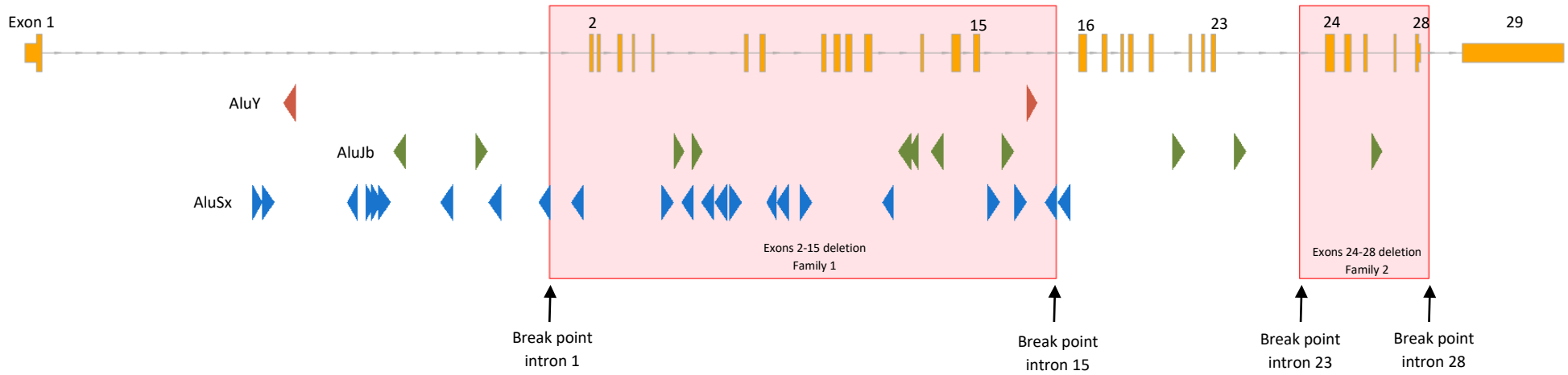

C.
Control of Patient A's run
2.0
1.5
1.0
0.5
0.0
Genes
MARS
KIF5A
TRPV4
Genomic position


D.
Control of Patient B's run
2.0
1.5
1.0
0.5
0.0
Genes
MARS
KIF5A
TRPV4
Genomic position


E.
Chr12:57,956,278
Chr12:57,968,335
AluSx
C T C C C A A A G T G C T G G G A T T A C A G G C G T G A G C C A C C G A
breakpoint region
(13 nt)
intron 1
intron 15


F.
Chr12:57,974,133
Chr12:57,977,210
A T G T G A A G G T A A T G A G T G C T T T C C G T T A G G C T A C A
breakpoint region
intron 23
intron 28


G.
Motor
Stalk
Tail
Family 1
c.(130-943_1717-533del)
p.(Gly44_Leu572del)
Family 2
c.(2539-605_*36+211del)
p.(Leu847_Ser1032delins33)

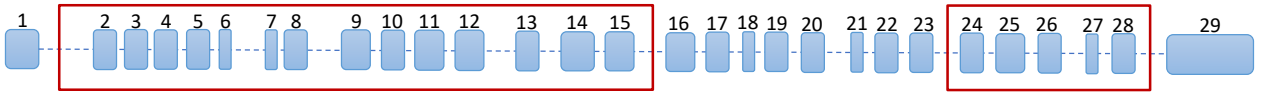

Exon 1
2
15
16
23
24
28
29
AluY
AluJb
AluSx
Exons 2-15 deletion
Family 1
Exons 24-28 deletion
Family 2
Break point
intron 1
Break point
intron 15
Break point
intron 23
Break point
intron 28

First Structural Variations in *KIF5A* responsible for Charcot-Marie-Tooth (CMT) disease discovered thanks to CovCopCan Software

Next Generation Sequencing

Bioinformatics analysis

First Structural Variations Detected in *KIF5A*



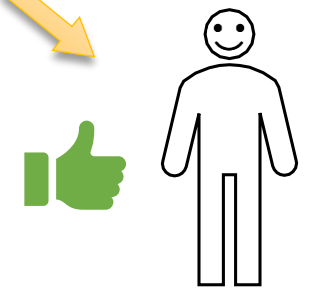
Deletion 1 Deletion 2

Genotype/Phenotype Correlation

Clinical examination

Symptoms	Patient A	Patient B	Patient C
CMT	✔	✔	✔
Pyramidal Sd	✔	✔	✔

Family 1 Family 2



Positive Diagnosis