Supplementary materials

eTable 1. Data about study subjects.

	Years	Referred to genetic counselling and DNA diagnostics (n)	Confirmed diagnosis (n)	Unique cases with diagnosis confirmed (n)
Medical Genetics clinic, Children's Clinical University Hospital, Riga, Latvia	2015-2020	137	62	62
Latvian Biomedical Research and Study Centre, Riga, Latvia	2008-2020	151	59*	30
Latvian Biomedical Research and Study Centre for **MD1**	2008-2020		8	8
Genera Ltd, Riga, Latvia for **MD1**	2008-2020		1	1
Medical Genetics laboratory Medical Genetics clinic, Children's Clinical University Hospital, Riga, Latvia for **SMA**	2008-2020	254	47	47
Scientific Laboratory of Molecular Genetics, Riga Stradins University, Riga, Latvia for **spinobulbar muscle atrophy**	2008-2020		5	5
Total				153

*21 patients overlap between Medical Genetics clinic, Children's Clinical University Hospital, Riga,

Latvia and Latvian Biomedical Research and Study Centre, Riga, Latvia due to the sequential

analysis, initiated by one center and transferred to another.

Allele frequency

The allelic frequency of variants in *CAPN3, CLCN1, HINT1, and SPG11* genes in the general population was determinated by direct sequencing. Each respective fragment was amplified using 10 pM of each primer (eTable 1, Metabion AG, Steinkirchen, Germany) and the HOT FIREPol polymerase master mix (Solis Biodyne, Cat # 04-25-0120) as well as 50 ng of the genomic DNA, followed by sequencing, using the BigDye™ Terminator v3.1 Cycle Sequencing Kit (ThermoFisher Scientific, Cat #4337455) and a ABI PRISM 3100 sequencer (Applied Biosystems, USA). All reactions were performed at standard conditions, according to manufacturer's recommendations.

The same protocol was used for reported variant validation.

SMN1/ SMN2 exon 7 copy number assay

SMN1, *SMN2*, and *RPP30* concentrations were simultaneously measured with a Bio-Rad QX200 Droplet Digital PCR system. ddPCR primers and probes were custom synthesized (IDT), as previously described [PMID:28826609]. The final concentration of primers and probes were 900 nmol/L and 250 nmol/L, respectively. For carrier testing, single well/ sample were used. ddPCR reaction assembly was as described in [PMID:28826609]. End-point PCR conditions were: 95 °C for 10 min, 40 cycles of denaturation at 94 °C for 30 s, annealing and extension at 58 °C for 1 min, and a final step at 98 °C for 10 min. QuantaSoft Analysis Pro (Bio-Rad) was used for droplet-cluster classification and application of Poisson function for absolute and relative *SMN1*, *SMN2* and *RPP30* copy numbers.

eTable 1. The primers used for PCR and direct sequencing

Gene	Amplicon size (bp)	Forward primer sequence 5'-3'	Reverse primer sequence 5'-3'
CAPN3 c.1746-20C>G	323	CGGGGATGGTGCTGACAT	CAGACTCCCATTCCCTCTG
HINT1 c.110G>C	377	TTCTTCCGAGCCTCTCCTC	CGGGGCAGATAACGAGTAAC
SPG11 c.2431C>T	593	AAGTGTGACAACCCCTTCATC	TCCACTCTTTTGGGGAACAC
CLCN1 c.2680C>T	220	CTTTCCCACTGCTCTTCAGC	GCAATCACATCCCCTGTTCC

CAG repeats in the AR gene

The CAG repeats in the AR gene were detected using capillary electrophoresis, and the results were confirmed by Sanger sequencing for two patients with the shortest CAG repeats, using the BigDye Terminator Kit v.3.1 (Thermo Fisher Scientific, USA), according to an

adapted manufacturer's protocol. The number of CAG repeats for longer alleles were calculated based on fragment size.

Statistical analysis plan by STROND guideline [1]

1
Title & abstract
Title and abstract 1
(a) Give the type of study design employed using a widely recognized term in the title or abstract - title
(b) The abstract should give an accurate summary of how the study was conducted and the main findings – abstract

Introduction
Background 2 Details of the scientific rationale for the study should be reported – page 3-4
Aims and objectives 3 State the specific aims and objectives of the study – page 4

Methods
Study design
4 Give a full description of the study design - Observational studies, retrospective studies
4a Give details of any study protocol (published or unpublished that gives additional useful information on the study design) - NA
4b If a pilot study has been conducted to inform the main study design, then the findings should be referenced – NA

Setting 5 Clearly defined (usually, but not always, on a geographic basis), and stable, with reliable information on in- and out-migration – page 4-5

Source population 6 Description of how all eligible members of the population were identified and through what data sources (e.g., hospitals, outpatient clinics, death certificates) – page 4
6a Source of data used for the study (e.g., administrative database, medical records). If administrative database used, algorithms for data extraction should be described – page 4
6b Description of the rate of hospital admission (if applicable) for the neurologic condition in the population - NA
6c Details of health care system in the country (study region) where the study was conducted (e.g., public vs private health care system) - page 4- 5
6d Description of how a person with the neurologic condition is referred (with the filters) in the country (study region) where the study was
Conducted – page 4-5
6e Description and characteristics of response rate/dropouts and exclusion rate if applicable NA
Participants 7 Definition of cases is clearly identified and presented in sufficient detail
7a Details of the sampling method are described (are participants representative of the source population?) - yes
7b Fully validated source of diagnosis or "reference-standard" criteria applied - yes

7c Definition and justification of the disease severity (preferably using a standardized severity scale) or staging of the disease - NA

7d Description of how types/subtypes of the neurologic disorder of interest are distinguished (if relevant) - NA

7e Description of how completeness of case ascertainment was assessed – all participating bodies involved

7f Description of whether completeness of case ascertainment was adequate – yes

Ethical approval 8 Details of ethics approval/informed consent/data governance should be reported – page 5

Measurement 9a Incidence studies

Give details of how incidence was determined (based on timing of data collection either prospectively or retrospectively) - NA

Definition and justification of timing of measurements - - NA

The data presented to some specified time period (usually whole years or person-time)- - NA

Raw numbers are reported in sufficient detail to calculate the appropriate rates (e.g., by age or sex) – NA

9b Prevalence studies

Give details of specific time points over which estimates are derived (usually defined as the number of cases existing in a specific time point) – pages 4-5

The data presented to some specified time period (usually whole years) – page 5

Raw numbers are reported in sufficient detail to calculate the appropriate rates (e.g., by age or sex) – page 5

9c

If disease burden is to be assessed, the study should report details of burden due to a variety of sources (e.g., disability, disability-adjusted life years, symptoms, financial, caregiver) - NA

9d

Report any arrangements for quality checks/data verification/ triangulation - NA

9e

Report details of the training of the person administering the instruments – PhD degree in Medecine

Statistical methods 10

If rates have been standardized (e.g., by age or sex), then the details of the standard population used should be given - NA

10a

If possible, 2 standard populations should be used, one with local relevance and the other to facilitate international comparisons – Table 2

10b

Description of any assumptions made in the calculations should be reported- NA

10c

An explanation of how missing data were addressed in the analyses - NA

10d

Provide a priori estimates of sample size/power assessment/precision of estimates assessment – page 4-5

10e

Description of any sensitivity analyses – NA

Results
Main findings 11

Consider a flow diagram that describes how participants were included in the study (useful in order to assess how a person with the neurologic condition of interest is referred [with the filters]) - NA
11a
Give appropriate rates with their associated 95% confidence intervals – page 7
11b
Report results of any sensitivity analyses – NA

Discussion
Key findings 12
Summarize the key findings in relation to the study aims and objectives – pages 8-10
Limitations 13
Discuss potential limitations of the study – page 10
13a
Include details of risk of bias (e.g., selection bias), completeness of case ascertainment, and data quality (assessment of its probability, size, and potential importance) – page 10
Interpretation 14
Interpret the results in the context of the evidence from other well- performed studies with similar designs and objectives – pages 8-10
14a
Reliability of the estimates (i.e., based on the reporting of the statistical methodology, and study design, measurement of key information) – Table 2
Generalizability 15
Discuss the external validity of the study findings – pages 8-10
15a
Are the results consistent with meta-analyses of descriptive epidemiologic studies on the same topic that cover different settings (if applicable)? - NA

References
1.   Bennett DA, Brayne C, Feigin VL, et al. Development of the Standards of Reporting of Neurological Disorders (STROND) checklist. *Neurology*. 2015;85(9):821-828. doi:10.1212/WNL.0000000000001866