Supplementary File

Methods

Patients

The Medical Ethical Committee of the University Medical Center Utrecht approved the study protocol. This study was registered at the Dutch registry for clinical studies and trials (http://www.ccmo-online.nl). We enrolled patients with SMA types 1-4 between September 2010 and August 2014 as part of a national study on SMA.[1] All patients were treatment-naïve for *SMN*-modulating therapies. Inclusion criteria were a genetically confirmed diagnosis of SMA. We used age at onset and acquired motor milestones for SMA classification; in case of discrepancy the latter defined the SMA type.[1, 2] SMA type 1 was defined as onset of symptoms before the age of 6 months and not acquiring the ability to sit unsupported, although some patients (type 1c) acquired other motor skills, such as head control or rolling from supine to prone or at least to one side at some stage in life. Patients with type 1c had previously been reported as surviving into adulthood with or without respiratory support.[1, 3-5] Patients with SMA type 2 showed first symptoms of disease between the ages of 6 and 18 months and learned to sit or even stand (but not walk) independently for a brief period.[1] Patients with the ability to sit unsupported as their highest acquired motor milestones, were classified as SMA type 2a, whereas patients who learned to stand or walk with support, even for a brief period of time, were classified as SMA type 2b.[1, 6] Patients with SMA type 3 had a disease onset after the age of 18 months and learned to walk independently. We used the sub-classification of SMA type 3a (disease onset before the age of 3 years) and type 3b (disease onset after the age of 3 years).[7] SMA type 4 (adult onset) was defined by an onset after 18 years of age in ambulatory patients.[2, 7]

Based on the phenotype and *SMN2* copy number, we classified each patient as concordant or discordant. We defined concordant patients as those who had an expected *SMN2* copy number

for their disease severity. We used this previously described model to define expected copy number: SMA type 1 has 2 copies of *SMN2*, type 2 has 3 copies of *SMN2*, and type 3 has 4 copies of *SMN2*.[8] We classified patients as discordant patients when disease severity milder or more severe then expected by the *SMN2* copy number.

Families were classified discordant when clinical phenotype and highest motor skills differed between siblings with the same *SMN2* copy number.

MLPA

DNA was extracted from peripheral blood. We used Multiplex Ligation-dependent Probe Amplification(MLPA) (MRC Holland www.mrcholland.nl) analysis to confirm the presence of homozygous deletions of exon 7 of the *SMN1* gene and to determine *SMN2* copy number. We performed the MLPA twice in every patient in separate experiments to confirm the *SMN2* copy number. In case of a hemizygous SMN1 deletion, we used Sanger sequencing to identify an additional point mutation in the second allele.

Sequence analysis

DNA was extracted from peripheral blood. We screened genomic DNA of all patients for variants in the genes: *Profilin 2*(*PFN2*; OMIM 176590; ENSG00000070087), *Fused in Sarcoma*(*FUS*; OMIM 137070; ENSG00000089280), *Transactive response DNA binding protein 43 kDa* (also called *TDP-43* or *TARDP*; OMIM 605078; ENSG00000120948) and *Plastin 3*(*PLS3*; OMIM 300131; ENSG00000102024) using multiplexed targeted resequencing, carried out on a high-throughput targeted next-generation sequencing platform (Miseq, Illumina). We applied DesignStudio (Illumina) to create a Truseq Custom Amplicon project for which the Standard Truseq Custom Amplicon Library preparation protocol was used.

We created bar-coded, paired-end sequencing libraries with 2x250 base pair read length per amplicon using prepared Truseq Custom Amplicon Kit (Illumina). We monitored the quality of separate sequencing runs with the help of Sequence Analysis Viewer Software (Illumina) and mapped sequencing reads to the human genome reference build GRCh37 using Burrows Wheeler Aligner (BWA 6.1). Base calling accuracy, measured by the Phred quality score (Q score) was presumed to be 'good' from a score of 30. The amplicons targeting coding, non-coding and 5'-3' untranslated regions of the *FUS, PFN2, TDP-43* and *PLS3* covered 96% of the regions of interest with good quality (quality score >30). Subsequent depth of coverage, quality filters, variant calling and variant annotation were performed using SAMtools v0.1.19, GATKv3.2 and the 1000 Genomes project. GATK filter criteria were relatively lenient, since we wanted to detect single exonic variants and their adjacent intronic regions in individual subjects ("QD" < 2.0; "FS" >500; "MQ" <40.0; "HaplotypeScore" <300.0; MQRankSum" <-12.5 and "ReadPosRankSum" <-8.0) that would later be validated using Sanger sequencing.

We performed Sanger sequencing to confirm the identified variants in the affected SMA patients. Analysis of variants was also performed in additional samples of patients with wild-type genotype and healthy controls to confirm the accuracy of the primers. Primer pairs were designed to fit the flanking sequence of the addressed variant or SNP using Primer3 software (Supplementary file Table S1).[9] PCR reactions were performed with HotstartTaq plus DNA polymerase (Qiagen, Alameda, CA, USA), 100ng genomic DNA (gDNA), 10 pmol of each primer and 250µM dNTPs (Roche, USA). We determined optimal annealing temperatures for each primer pair by Tgradient on an iCycler PCR (Bio-Rad, Hercules, CA, USA) (Table S1). Amplicons were visualised on agarose gels, and sequenced in both directions using the ABI Prism BigDye Terminator Cycle Sequencing V3.1 kit (Applied Biosystems, Foster City, CA, USA) on an ABI Prism 3730 sequencer (Applied Biosystems, Foster City, CA, USA). Sequences were assembled to a consensus sequence and evaluated for the predicted SNPs

using the DNA Baser sequence analysis software (DNA Sequence Assembler v4 (2013), Heracle BioSoft).

PLS3 Droplet Digital PCR (ddPCR) analysis

We used PAXgene blood RNA tubes (BD Biosciences, San Jose, CA, USA) for storage and stabilization of RNA from peripheral blood. RNeasy Mini Kit (Qiagen, Dusseldorf, Germany) was used to extract messenger-RNA from blood. RNA was DNase-digested with TURBO DNA-free kit (Ambion). RNA concentration was determined by spectrophotometer absorbance determination and quality assessed by nanodrop (NanoDrop 2000, Thermo Scientific) analysis (absorbance of 230, 260 and 280nm). We defined a ratio (260/280) of ±2.0 as 'pure'. Quality and integrity control of PAXgene samples was performed with an Agilent 2200 TapeStation. We used a RNA Integrity Number (RIN) cut-off value above 5.6. A High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, No.4368814) was applied for the reverse transcription of 150ng RNA to cDNA.

We used 3 commercially available assays for PLS3, TBP and HRPT1(Assay numbers: PLS3= Hs00418605_g1; TBP= dHsaCPE505863; HPRT1= dHsaCPE5192872) for droplet digital PCR(ddPCR) analysis using QX200™ Droplet Digital PCR System (Bio-Rad, Hercules, CA, USA). The PLS3 assay (Life Technologies), TPB assay and HPRT1 assay (Bio-Rad, Hercules, CA, USA) were validated by a temperature gradient on control cDNA.[10] In brief, 22µl reactions contained 1µl of cDNA, 1 µl 20X of reference target probe mix (TBP / HPRT1, HEX-labeled), 1 µl of 20X target probe mix (PLS3, FAM-labeled), 11µl of 2X ddPCR Supermix for probes (No dUTP) and 8 µl RNase-/DNase free water. We mixed 20µl of the reaction mix with 70µl of Droplet Generation Oil (Bio-Rad) and partitioned it into droplets in Automated Droplet Generator (Bio-Rad). PCR amplification for PLS3 in

combination with TBP and HPRT1 reference genes was performed in the Bio-Rad T100 thermal cycler (10 min at 95°C; 40 cycles of 30 sec at 95°C and 60 sec at 56,1°C; and 1 cycle of 10 min at 98°C, followed by 12°C for cooling with a 2°C/s ramp rate). After amplification, we analyzed the droplets in a QX200 droplet reader (droplet flow cytometer) (Bio-Rad, Hercules, CA, USA). mRNA concentrations were calculated in copies per 1 ng cDNA.

TPB and HRPT1 levels were both used as a reference. The final expression level of PLS3 was calculated using the geometric mean of two separate experiments.[11]

Statistical analysis

We analysed the effect of the newly detected missense variants on protein structure or function with four prediction programs: PolyPhen[12] (http://genetics.bwh.harvard.edu/pph2/), SIFT[13] (ttp://sift.bii.a-star.edu.sg), Panther[14] (http://www.pantherdb.org) and ExAC Browser[15] (http://exac.broadinstitute.org). We evaluated variants in non-coding regions for disruption or creation of exonic splicing enhancer sites with Netgene2[16] (http://www.cbs.dtu.dk/services/NetGene2/) and Human Splice Finder[17] (http://www.umd.be/HSF3/). We determined minor allele frequencies with the help of ExAC Browser[15] (http://exac.broadinstitute.org), gnomAD[15] (http://gnomad.broadinstitute.org) and Project MinE databrowser[18] (http://databrowser.projectmine.com) databases. We used no restrictions on allele frequencies in our analysis.

We analysed the effects of expression levels with the help of mean, medians and SD for continuous variables and proportions for categorical variables. Correlations were analyzed using the Spearman's rank correlation coefficient. Univariate and multivariate tests including dichotomous data were performed using logistic regression. Comparison of data between SMA types, *SMN2* copy numbers and between patients and controls was performed using Kruskal-Wallis(KW) test, Mann-Whitney U(MW) test or Chi-square analysis. P-values <0.05

were considered statistically significant. P-value of <0.05 was used in case of Bonferroni correction for multiple testing.

We used SPSS (IBM SPSS Statistics version 23.0.0.2 (Mac), Inc., Chicago, IL) for statistical analysis.

References

1. Wadman RI, Stam M, Gijzen M, et al. Association of motor milestones, SMN2 copy and outcome in spinal muscular atrophy types 0-4. J Neurol Neurosurg Psychiatry 2017.

2. Mercuri E, Bertini E, Iannaccone ST. Childhood spinal muscular atrophy: controversies and challenges. Lancet Neurol 2012;11:443-452.

3. Cobben JM, Lemmink HH, Snoeck I, Barth PA, van der Lee JH, de Visser M. Survival in SMA type I: a prospective analysis of 34 consecutive cases. Neuromuscul Disord 2008;18:541-544.

4. Rudnik-Schoneborn S, Berg C, Zerres K, et al. Genotype-phenotype studies in infantile spinal muscular atrophy (SMA) type I in Germany: implications for clinical trials and genetic counselling. Clin Genet 2009;76:168-178.

5. Petit F, Cuisset JM, Rouaix-Emery N, et al. Insights into genotype-phenotype correlations in spinal muscular atrophy: a retrospective study of 103 patients. Muscle Nerve 2011;43:26-30.

6. Wadman RI, Wijngaarde CA, Stam M, et al. Muscle strength and motor function throughout life in a cross-sectional cohort of 180 patients with spinal muscular atrophy types 1c-4. Eur J Neurol 2018;25:512-518.

7. Zerres K, Rudnik-Schoneborn S. Natural history in proximal spinal muscular atrophy. Clinical analysis of 445 patients and suggestions for a modification of existing classifications. Arch Neurol 1995;52:518-523.

8. Ruhno C, McGovern VL, Avenarius MR, et al. Complete sequencing of the SMN2 gene in SMA patients detects SMN gene deletion junctions and variants in SMN2 that modify the SMA phenotype. Hum Genet 2019.

9. Untergasser A, Cutcutache I, Koressaar T, et al. Primer3--new capabilities and interfaces. Nucleic Acids Res 2012;40:e115.

10. Wadman RI, Stam M, Jansen MD, et al. A Comparative Study of SMN Protein and mRNA in Blood and Fibroblasts in Patients with Spinal Muscular Atrophy and Healthy Controls. PLoS One 2016;11:e0167087.

11. Vandesompele J, De Preter K, Pattyn F, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 2002;3:RESEARCH0034.

12. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010;7:248-249.

13. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. Nat Protoc 2016;11:1-9.

14. Mi H, Muruganujan A, Thomas PD. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res 2013;41:D377-386.

15. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;536:285-291.

16. Brunak S, Engelbrecht J, Knudsen S. Prediction of human mRNA donor and acceptor sites from the DNA sequence. J Mol Biol 1991;220:49-65.

17. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res 2009;37:e67.

18. Project Min EALSSC. Project MinE: study design and pilot analyses of a large-scale whole-genome sequencing study in amyotrophic lateral sclerosis. Eur J Hum Genet 2018;26:1537-1546.