

Research Report

Closing the Gap - Detection of 5q-Spinal Muscular Atrophy by Short-Read Next-Generation Sequencing and Unexpected Results in a Diagnostic Patient Cohort

Stephanie Kleinle^{a,1}, Veronika Scholz^{a,1}, Anna Benet-Pagés^{a,b,1}, Tobias Wohlfrom^a, Stefanie Gehling^a, Florentine Scharf^a, Simone Rost^a, Eva-Christina Prott^c, Susanne Grinzinger^d, Anna Hotter^e, Verena Haug^f, Sabine Niemeier^g, Lucia Wiethoff-Ubrig^h, Tim Hagenackerⁱ, Klaus Goldhahn^j, Arpad von Moers^j, Maggie C. Walter^k, Peter Reilich^k, Katja Eggermann^l, Florian Kraft^l, Ingo Kurth^l, Hannes Erdmann^{a,k}, Elke Holinski-Feder^{a,m}, Teresa Neuhann^{a,*} and Angela Abicht^{a,k,*}

^aMedical Genetics Center, Munich, Germany

^bInstitute of Neurogenomics, Helmholtz Center Munich, Neuherberg, Germany

^cPractice for Human Genetic, Wuppertal, Germany

^dChristian Doppler Clinic, Neurology, University Hospital Salzburg, Salzburg, Austria

^eDepartment of Neurology, Medical University of Innsbruck, Innsbruck, Austria

^fNeuropediatrics, University Medical Center Mainz, Mainz, Germany

^gWestbrandenburg Clinic, Neuropediatrics, Potsdam, Germany

^hChildren's and Adolescents' Hospital Datteln, Neuropediatrics, Witten/Herdecke University, Datteln, Germany

ⁱDepartment of Neurology and Center for Translational Neuro- and Behavioral Sciences (C-TNBS), University Hospital Essen, Essen, Germany

^jDepartment of Pediatrics and Neuropediatrics, DRK Clinics Berlin, Berlin, Germany

^kFriedrich Baur Institute at the Department of Neurology, University Hospital, LMU Munich, Munich, Germany

^lInstitute for Human Genetics and Genomic Medicine, Medical Faculty, RWTH Aachen University, Aachen, Germany

^mDepartment of Medicine IV, University Hospital, LMU Munich, Munich, Germany

Accepted 20 June 2023

Pre-press 5 July 2023

Published 8 September 2023

Abstract.

Background: The importance of early diagnosis of 5q-Spinal muscular atrophy (5q-SMA) has heightened as early intervention can significantly improve clinical outcomes. In 96% of cases, 5q-SMA is caused by a homozygous deletion of *SMN1*. Around 4 % of patients carry a *SMN1* deletion and a single-nucleotide variant (SNV) on the other allele. Traditionally, diagnosis is based on multiplex ligation probe amplification (MLPA) to detect homozygous or heterozygous exon 7 deletions in *SMN1*. Due to high homologies within the *SMN1/SMN2* locus, sequence analysis to identify SNVs of the *SMN1* gene is unreliable by standard Sanger or short-read next-generation sequencing (srNGS) methods.

¹These authors contributed equally.

*Correspondence to: Angela Abicht, Medical Genetics Center, Bayerstr. 3–5, 80335 Munich, Germany. Tel.: +49 89 30 90

886 0; Fax: +49 89 30 90 886 66; E-mail: angela.abicht@mgz-muenchen.de.

Objective: The objective was to overcome the limitations in high-throughput srNGS with the aim of providing SMA patients with a fast and reliable diagnosis to enable their timely therapy.

Methods: A bioinformatics workflow to detect homozygous *SMN1* deletions and *SMN1* SNVs on srNGS analysis was applied to diagnostic whole exome and panel testing for suggested neuromuscular disorders (1684 patients) and to fetal samples in prenatal diagnostics (260 patients). SNVs were detected by aligning sequencing reads from *SMN1* and *SMN2* to an *SMN1* reference sequence. Homozygous *SMN1* deletions were identified by filtering sequence reads for the „gene-determining variant“ (GDV).

Results: 10 patients were diagnosed with 5q-SMA based on (i) *SMN1* deletion and hemizygous SNV (2 patients), (ii) homozygous *SMN1* deletion (6 patients), and (iii) compound heterozygous SNVs in *SMN1* (2 patients).

Conclusions: Applying our workflow in srNGS-based panel and whole exome sequencing (WES) is crucial in a clinical laboratory, as otherwise patients with an atypical clinical presentation initially not suspected to suffer from SMA remain undiagnosed.

Keywords: 5q-SMA, spinal muscular atrophy, clinical genetics, neuromuscular disorders, bioinformatics, dark genes

INTRODUCTION

5q-SMA is one of the most common autosomal recessive inherited disorders with a frequency of approximately 1:7,000 among newborns in Germany and 1:10,000 worldwide, resulting in a carrier frequency of 1:40 in Germany and 1:50 worldwide [1–3]. The disease is characterized predominantly by degeneration of alpha-motoneurons in the spinal cord. Patients present with muscular hypotonia, as well as proximal weakness and atrophy, especially of the lower extremities. According to the maximal motor abilities acquired (never-sitting, sitting, walking) and onset of the disease, SMA patients were originally classified into five types (0, 1, 2, 3, and 4) [4]. As a result of new therapeutic options, a new classification was implemented that better reflects the continuum spectrum of the disease and now divides patients into non-sitters, sitters, and walkers [5, 6].

The disease-underlying gene is the survival motor neuron 1 (*SMN1*) gene on chromosome 5q13.2. *SMN1* has an almost identical but functionally far less relevant copy gene, *SMN2*. Whereas *SMN1* exclusively produces correct RNA and protein, *SMN2* predominantly produces mis-spliced RNA and non-functional protein. Only about 10% of transcripts are correctly spliced and generate a protein identical to the *SMN1* encoded protein [1]. Both genes, *SMN1* and *SMN2*, are located in an approximately 500 kb segment of a copy number variant (CNV). This CNV is highly polymorphic in structure, length, and orientation in the population. Affected individuals with 5q-SMA usually lack *SMN1* and carry between one and six *SMN2* copies per genome, affecting the severity of the phenotype [3].

In 96% of SMA patients, the disease is caused by a homozygous deletion of exons 7 and 8 or only of exon 7, as a result of either a deletion or a *SMN1*-*SMN2* gene conversion event. In approximately 4% of 5q-SMA patients, compound heterozygosity is present, namely – a *SMN1* deletion CNV on one allele and a pathogenic *SMN1* SNV on the second allele [7]. Due to high homologies within the *SMN1*/*SMN2* locus on chromosome 5q13.2, analysis of the *SMN1* gene is not possible by srNGS methods. Such regions of the genome, where short reads cannot be mapped unambiguously due to sequence homology, are called “dark regions” or “dead zones” for NGS technology [8, 9]. Therefore, the standard diagnostic workup in patients with clinically suspected SMA is based on the detection of a homozygous exon 7 *SMN1* deletion by multiplex ligation probe amplification (MLPA). However, patients with an atypical clinical presentation remain undiagnosed if only srNGS exome analysis is applied.

Within this study, we developed and implemented a specific high-throughput bioinformatics workflow for the *SMN1/2* locus that masked the *SMN2* gene to detect homozygous *SMN1* deletions and *SMN1* SNVs on srNGS data. In two patients with typical symptoms of SMA and a known heterozygous *SMN1* deletion, we were able to identify heteroallelic pathogenic SNVs confirming the diagnosis of 5q-SMA. Applying our workflow to patients that underwent short-read panel or exome sequencing in a clinical laboratory, we identified homozygous *SMN1* deletions in six patients with atypical clinical presentations initially not suspected to suffer from SMA. Moreover, we were able to identify two patients with each two heterozygous pathogenic SNVs.

MATERIALS AND METHODS

Patient cohorts

We retrospectively analyzed two cohorts of patients who underwent diagnostic srNGS-based exome-analysis within a 12-months period between November 2021 and November 2022. Cohort 1 consisted of 1684 patients with a neuromuscular phenotype as evidenced by the presence of at least one of the Human Phenotype Ontology (HPO) [10] phenotypes or its direct descendants: abnormality of the musculature and peripheral neuropathy. Cohort 2 consisted of 260 patients (fetal samples) that underwent exome sequencing as part of prenatal diagnostics within the one-year period. The medical indication of prenatal diagnostics was the presence of fetal ultrasound abnormalities.

As positive and negative controls to test the accuracy of the 5q-SMA variant detection workflow, we used data from 212 patients analyzed in parallel with SMA-MLPA and srNGS.

Compliance with ethical standards

Written informed consent was obtained from each study participant. As part of the informed consent process prior to analysis, patients were educated about the significance and limitations of NGS diagnostics and the potential for incidental findings. In all analyses, consent for the report of incidental findings with therapeutic relevance was given. All genetic analyses and investigations were performed in accordance with the guidelines of the Declaration of Helsinki and were approved by local institutions (Bayerische Landesärztekammer, 2019-210).

Short-Read Next Generation Sequencing

Short-Read Next Generation Sequencing (Twist Human Comprehensive Exome + Mitochondrial Genome) comprising 19182 genes, respectively, was performed on genomic DNA samples using an Illumina NovaSeq 6000 system. NGS was carried out as 150 bp paired-end runs using v2.0 sequencing by synthesis (SBS) chemistry. Processing of NGS raw data and variant interpretation was done with the varvis® (Limbus Medical Technologies GmbH, Rostock) genomics platform.

Sequence variants were classified according to the ACMG (American College of Medical Genetics and genomics) guidelines [11] using the 5-tier classifi-

cation system: class 5 (pathogenic), class 4 (likely pathogenic), class 3 (uncertain variants or variants of unknown significance, VUS), class 2 (likely benign) and class 1 (benign).

5q-SMA variant detection workflow

Using (1) a *SMN2*-masked pipeline and (2) a specific *SMN1/2*-variant filtering, we set up a bioinformatics workflow for the *SMN1/2* locus to detect (i) a homozygous *SMN1* deletion or (ii) homozygous, hemizygous or heterozygous *SMN1* SNVs and indels from srNGS data.

1. Bioinformatics SMN2-masked pipeline

Due to high homologies within the *SMN1/SMN2* locus, short reads cannot be unambiguously mapped. If a sequence read aligns to two or more regions equally well, most aligners randomly map the read to one region and assign a low mapping quality (MAPQ=0 for BWA) [12]. These reads are excluded by the variant callers making correct variant detection of the *SMN1* gene impossible with a standard bioinformatics pipeline. A technique to generate unambiguous mappings is to hide (\rightarrow mask) homologous regions except for one, so that all reads are mapped to that region. For *SMN1*, the paralogous regions comprising *SMN2* (chr5:69,345,100 - 69,373,700) were “masked” in the CRGh37/hg19 reference genome (converting the DNA sequence bases into Ns) to force NGS reads to map to a single location in a secondary alignment step (Fig. 1A) [8]. This is a simplification for the mapper with the implication that it remains unknown if the reads actually originate from the hidden homologous sequence or not. It has to be considered that the assumed diploid sequence for the variant calling is not anymore true and is at least doubled or even more tripled resulting in variant allele frequencies (VAF) \neq 0.5 for heterozygous variants. Variant calling was done with the vardict caller, a specific caller for low frequency variants [13]. Sequence variants (i.e., SNVs and indels) were annotated and prioritized using the varvis® (Limbus, Rostock) tool.

2) SMN1/2 variant filtering workflow

After running the bioinformatics *SMN2*-masked pipeline a specific *SMN1/2*-variant filtering was applied for:

- i. Detection or exclusion of *SMN1* deletions (Fig. 1B): A homozygous deletion of exon 7 of *SMN1* was detected by identi-

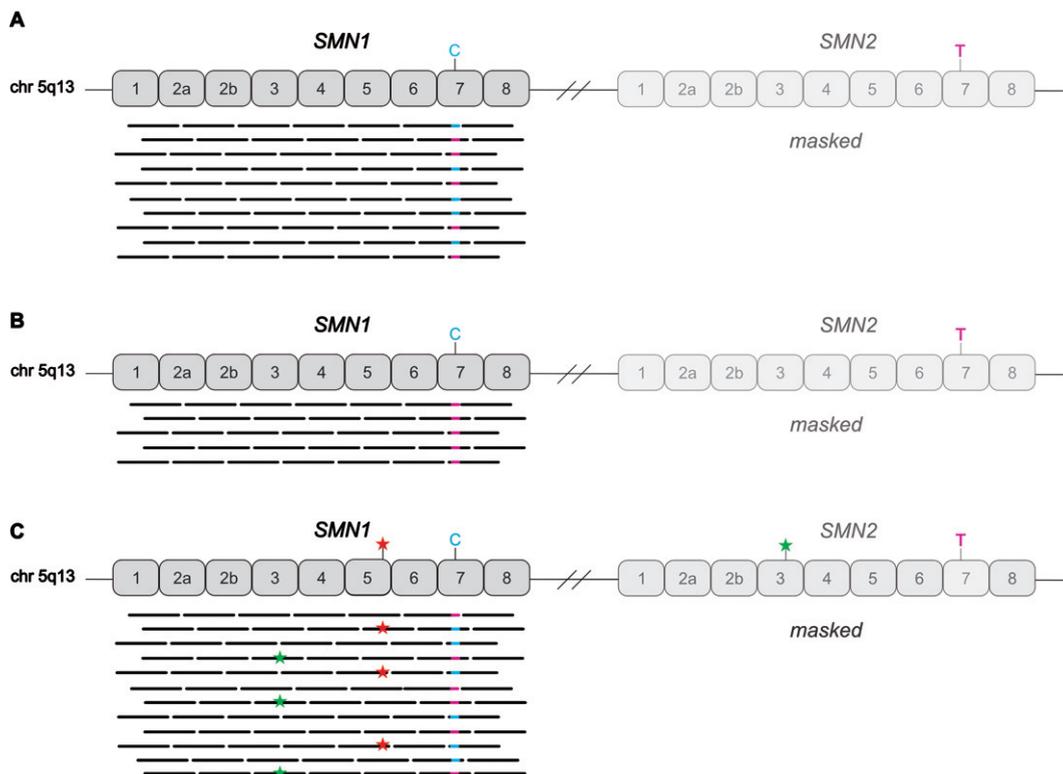


Fig. 1. Schematic representation of the *SMN1/2* locus on chromosome 5q. *SMN1* and *SMN2* genes differ in the “Gene-determining variant” (GDV) NM_000344.4:c.840C>T p.(Phe280=) (rs1164325688) in exon 7, depicted as C (blue) in *SMN1* and T (red) in *SMN2*. By masking of *SMN2* in the reference genome, all *SMN1/2* reads are mapped to *SMN1*. *SMN1* can be analyzed for SNVs by the caller. (A) Example of a normal control: The T (GDV of *SMN2*) is called in a VAF of < 0.9 indicating absence of a homozygous deletion of exon 7 of *SMN1*. (B) Example of a patient with a homozygous *SMN1* deletion: There are no reads with the C (*SMN1*). Thus, the variant caller indicates T (GDV of *SMN2*) with a VAF of > 0.9 (“homozygous”). (C) Example of a patient with a pathogenic SNV in *SMN1* (indicated as red star) and a benign variant in *SMN2* (indicated as green star). After applying the *SMN2*-masked pipeline, SNVs in *SMN1* as well *SMN2* reads are called. It has to be considered that the assumed diploid sequence for the variant calling is not anymore true and is at least doubled or even more tripled resulting in variant allele frequencies (VAF) \neq 0.5 for heterozygous variants. The VAF depends on the *SMN1* and *SMN2* copy number of the individual. A heterozygous SNV in one *SMN1* copy may have a low VAF (e.g. of 0.2 in case of two *SMN1* and three *SMN2* copies).

fyng homozygous SNV calls (VAF > 0.9) of the „gene-determining variant“ (GDV) NM_000344.4:c.840 C > T p.(Phe280 =) (rs1164325688) which is unique to *SMN2*. Homozygosity for the GDV c.840 C > T indicates the absence of *SMN1* copies and therefore the presence of a homozygous *SMN1* deletion. By contrast, heterozygous SNV calls (VAF < 0.9) indicate the presence of at least one *SMN1* copy and therefore exclude a homozygous *SMN1* deletion.

- ii. Detection of homozygous, heterozygous, or hemizygous SNVs and indels (Fig. 1C, Fig. 2): all SNVs from the *SMN1/SMN2* combined reads are called as apparent SNVs in *SMN1*. Considering that the assumed diploid sequence for variant calling is no longer correct, variants

with allele frequencies different from 0.5 must be regarded as potential disease-causing variants as their true zygosity referred to *SMN1* might be heterozygous, homozygous, or hemizygous. In patients with significant pathogenic variants (ACMG class 4 or 5), additional MLPA analysis was performed to detect or exclude heterozygous *SMN1* deletion and to estimate *SMN2* copy number (CN).

Verification of the accuracy of the 5q-SMA variant detection workflow

To assess the accuracy of detecting or excluding homozygous *SMN1* deletions by our 5q-SMA variant detection workflow, we evaluated data from 212 patients with known SMA-MLPA results and data

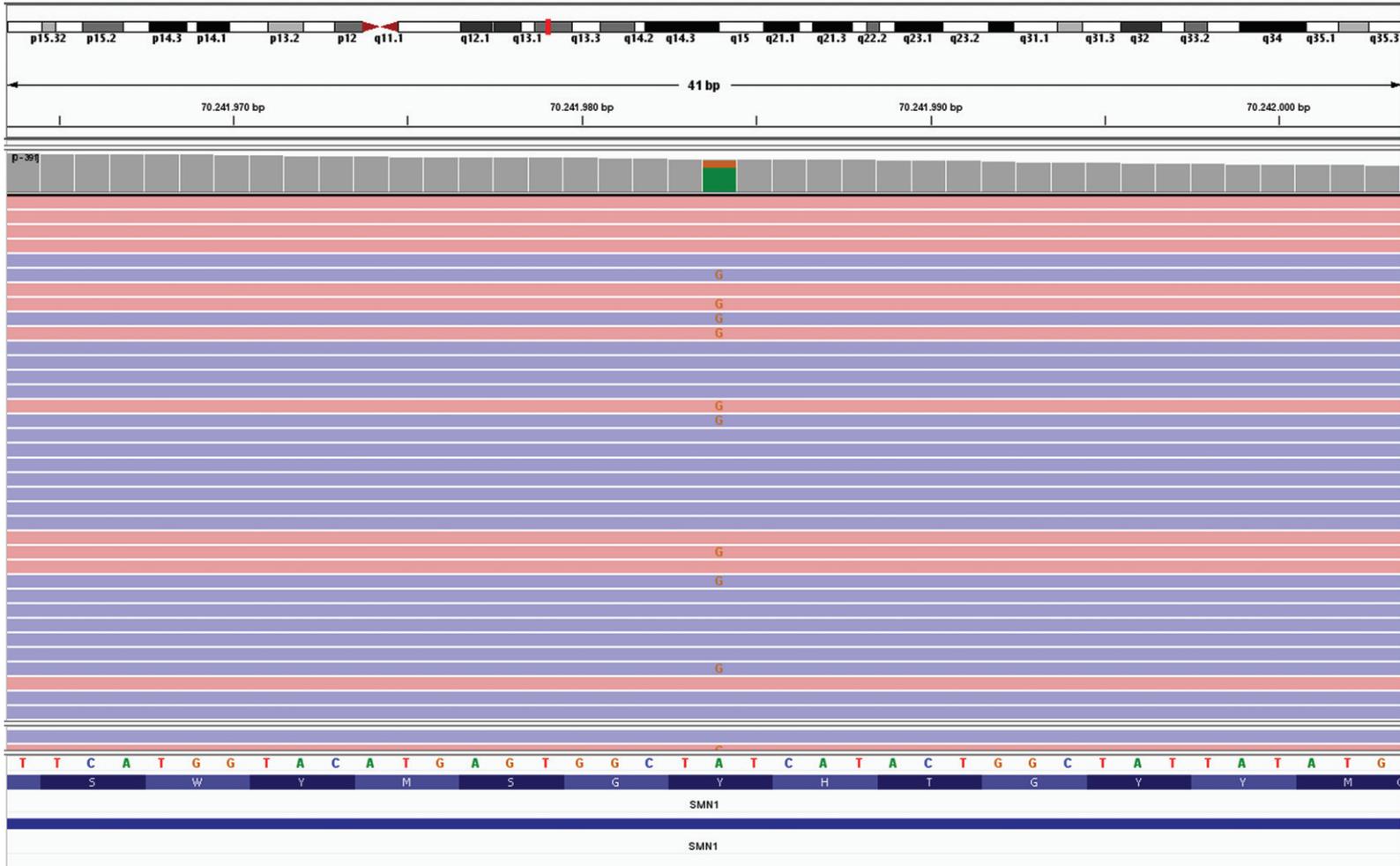


Fig. 2. Example of a heterozygous pathogenic variant in *SMN1*, NM_000344.4:c.815A>G p.(Tyr272Cys) in the IGV (Integrative Genomics Viewer (27)).

from srNGS. In addition, we verified the accuracy of SNV calling by analyzing the Giab *SMN1* benchmarking reference sample HG002 and by evaluating the called VAF of three common population SNVs [NM_000344.4:c.462A>G p.Gln154Gln (rs4915), NM_000344.4:c.835-44 G>A (rs1454173648), and NM_000344.4:c.84 C>T p.Ser28=(rs1554081948)] located in the *SMN1/2* dark region.

Multiplex Ligation-dependent Probe Amplification (MLPA)

Patients with *SMN1* deletions and/or pathogenic *SMN1* variants were subsequently analyzed by MLPA to confirm the *SMN1* and determine the *SMN2* CN. Genomic DNA was analyzed with the Salsa MLPA SMA P021 kit (MRC-Holland, Amsterdam, The Netherlands) according to the manufacturer's instructions. The MLPA products were separated by capillary electrophoresis using the ABI-3130XL Genetic Analyzer (Applied Biosystems, Foster City, CA). The results were imported into GeneMarker software (Softgenetics, State College, CA) for the MLPA data analysis using internal control probe normalization with the MLPA ratio as analysis method and peak area as quantification method.

Long-range PCR-based Nanopore sequencing for haplotype determination

DNA was isolated from blood using a salting-out method. The *SMN1/2* sequence was amplified by three long-range(LR) PCRs, with two LR-PCRs with ~13.5kb size spanning the whole and the third LR-PCR the middle of the genes [14]. LR-PCR was carried out with the AccuPrime Taq DNA Polymerase (Thermo Fisher Scientific, Waltham, Massachusetts, USA). PCR products were pooled and cleaned-up with 0.4x CleanNA CleanNGS beads (CleanNA, Waddinxveen, Netherlands). Nanopore sequencing was carried out according to the manufacturer's protocol, using the EXP-NBD104 together with the SQK-LSK109 on a FLO-FLG001 flowcell (Oxford Nanopore Technologies, Oxford, UK) [15].

Base calling was carried out with guppy (6.4.2). The resulting FASTQ files were filtered with NanoFilt (2.8.0) regarding read length (>12kb) and quality (>8), aligned against a *SMN2* masked GRCh38 reference genome using minimap2 (2.24) and converted, sorted, and indexed with SAMtools (1.16). IGV (2.15.2) was used for data visualization. Variant calling was done with Medaka (1.5.0) and WhatsHap

(1.3) to split *SMN1* and *SMN2* reads and generate VCF files for both genes.

Genome Nanopore Sequencing

Genomic DNA was purified from peripheral blood samples using the NucleoMag Blood 3 mL Kit (Macherey-Nagel), and 1 µg gDNA was subjected to sequencing. DNA end-prep was performed using the NEBNext Ultra II End Repair/dA-Tailing Module (E7546, New England Biolabs) followed by adapter ligation with the NEBNext Quick Ligation Module (E6056, New England Biolabs). Sequencing libraries were prepared with the Ligation Sequencing Kit 1D (SQK-LSK109, Oxford Nanopore Technologies) and run on R9 flow cells on the Oxford Nanopore Technology GridION instrument. DNA purifications at each step of library preparation were performed using Agencourt AMPure XP beads (A63882, Beckman Coulter). Sequencing was performed to a final mean sequencing depth of 7.13 x. Structural variant (SV) calling was performed with Nanovar V.1.3.8 [16] using Minimap2 (2.17-r941) [17], blast+ (2.10.1)22, and hs-blastn (v0.0.5) [18]. The human reference sequence was GRCh38/hg38. The final structure and exact breakpoint information was deduced from a combination Nanovar information and manual inspection of soft clipped read sequences at predicted breakpoints using the Integrative Genomics Viewer V2.9.024.

RESULTS

The 5q-SMA variant detection workflow was verified to be accurate in the detection of homozygous *SMN1* deletions. All homozygous *SMN1* deletions in eleven positive controls with known MLPA results were detected by a c.840 C>T VAF from 0.995 to 1.0 (Suppl. Table 1). In contrast, all 190 negative controls with a *SMN1* exon 7 CN of 2 or more revealed a VAF from 0.190 to 0.610 thus indicating absence of a homozygous *SMN1* deletion (Suppl. Table 1). The accuracy of SNV calling was confirmed by analyzing seven individuals with different CNs of *SMN1* and *SMN2*, and the Giab *SMN1* benchmarking reference sample HG002 [23] for three common population SNVs located in the *SMN1/2* dark region (Suppl. Table 2). We were able to call all SNVs at a VAF in line with the expected proportion dependent on *SMN1/SMN2* CN in all samples. The application of our newly implemented 5q-SMA

variant detection workflow (Fig. 1 and 2) to a diagnostic patient cohort revealed a diagnosis of 5q-SMA in nine out of 1684 patients with neuromuscular phenotypes (cohort 1). One patient with 5q-SMA was identified in a cohort of 260 prenatal samples (cohort 2).

Patients with heterozygous SMN1-deletion and SNV

In two neuromuscular patients (ID1, ID2) with a previous MLPA result indicating a heterozygous deletion of exon 7 of *SMN1*, the *SMN2*-masked bioinformatics workflow identified the heteroallelic pathogenic SNV c.821 C>T p.(Thr274Ile) (ACMG class 5). MLPA revealed that both patients carried two *SMN2* copies each. Both unrelated patients were adults (ID1:44 years of age and ID2:18 years of age) with a clinical diagnosis of SMA that prompted genetic testing for 5q-SMA (Table 1).

Patients with homozygous SMN1-deletions

Homozygous *SMN1* deletions were detected in five neuromuscular patients, initially not suspected of being affected by 5q-SMA (ID3, ID4, ID5, ID7, ID8). Diagnostic tests for various other suspected neuromuscular diseases were requested in these patients (Table 1). At the time of diagnosis, these five patients were between eight and 47 years old. While one patient (17 years at diagnosis) had three *SMN2* copies, the other four patients were found to have four *SMN2* copies.

One homozygous *SMN1* deletion was identified in a retrospective NGS data analysis of a fetal sample (ID6). The clinically severely affected newborn child had a positive result in newborn screening for SMA and turned out to have a single *SMN2* copy.

Patients with two heterozygous SNVs

In two patients (ID9, ID10), we identified compound heterozygous SNVs, which were classified as (likely) pathogenic according to ACMG guidelines (Table 2). Both patients carried a heterozygous truncating variant in combination with a heterozygous missense variant. One patient was a 48-year-old female (ID9) with one *SMN2* copy and SNVs c.469 C>T p.(Gln157*) and c.821 C>T p.(Thr274Ile). Long-read sequencing (long-range PCR-based Nanopore sequencing) showed that the sequence variants are on different *SMN1* alleles,

thereby confirming compound heterozygosity and the diagnosis of 5q-SMA. The second patient (ID10) was a four-month-old infant with two *SMN2* copies and SNVs c.549del p.(Lys184Serfs*29) and c.815A>G p.(Tyr272Cys). The combined heterozygosity of the latter variants was confirmed by analysis of parental samples. In addition, long-read sequencing (Genome Nanopore Sequencing) of parental blood samples confirmed that both variants are located in *SMN1* and not *SMN2*. In both the infant and adult patients, a previously suggested diagnosis of 5q-SMA was “excluded by MLPA”.

DISCUSSION

Since approved molecular therapies become available for SMA, this diagnosis has become one of the diseases with the greatest therapeutic relevance in the field of neuromuscular diseases. This is also reflected in the fact that it has been included in newborn screening in Germany, among other countries world-wide [19], since treatment is most successful when therapeutic intervention is provided as early as possible.

Although SMA is one of the most common autosomal recessive disorders, genetic diagnosis still faces challenges. The most common pathomechanism, homozygous deletion of exon 7 of the *SMN1* gene which underlies 96% of SMA patients, is analyzed by genotyping PCR techniques or MLPA, and less commonly by quantitative PCR techniques. While MLPA detects homozygous as well as heterozygous *SMN1* deletions, not all quantitative PCR techniques are validated to identify heterozygous deletions which are present in about 4% of patients heteroallelic to pathogenic sequence variants in *SMN1*.

However, this targeted testing approach requires the clinician to explicitly recognize the disease as a suspected diagnosis. In contrast to many other neuromuscular diseases, 5q-SMA cannot be detected by short-read NGS-based panel or exome analysis using standard bioinformatics pipelines. This is because the *SMN1/SMN2* locus on chromosome 5q13.2 belongs to the “dark regions” or “dead zones” for NGS technology [8, 9] where genomic complexity with variable copy-numbers and high homologies do not allow assignment of reads to either *SMN1* or *SMN2*. As a result, the detection of SNVs and hetero- or homozygous deletions of the *SMN1* gene is not covered by standard srNGS protocols.

In order to close this crucial diagnostic gap and to allow the relevant diagnosis of 5q-SMA also in

Table 1
Patients diagnosed with SMA based on the implemented 5q-SMA variant detection workflow

Patient ID	Suspicion of SMA when NGS analysis is initiated:	phenotype/short clinical history/previous diagnostic workup
ID1	Yes	Symptom-onset with 6 months, sitting without support at 7 months, gained ability to stand with 10 months, lost it with 3 years, never able to walk without assistance. Affected brother with similar severity.
ID2	Yes	Walking was possible until age 4 years, wheelchair-bound since age 4 years, standing until age 9 years, neuropathic scoliosis, scoliosis surgery at the age of 15 years, contractures of elbow, hip, knee and ankle joints, no reflexes, proximal and distal tetraparesis, arms can be lifted to the mouth when placed on a table.
ID3	No	First symptoms at 6 months of age. Walking at 15 months of age, persistent tip-toeing. At 5 months of age difficulties in climbing stairs and in running, exercise intolerance, suspected episodic ataxia. Muscle biopsy (left thigh) showed progressive neurogenic muscle atrophy with predominance of type 2 fibers. Indication for diagnostic testing (exome): Suspicion of axonal neuropathy or myopathy.
ID4	No	Always active in sports, at around 30 years of age mild difficulty in standing up after falling while playing soccer (retrospective observation); after lumbar disc surgery at 44 years of age difficulty in gaining strength during rehabilitation and awareness of proximally pronounced tetraparesis and CK elevation, muscle biopsy inconclusive, EMG with marked neurogenic changes leading to indication for diagnostic testing (neuromuscular panel): Amyotrophic lateral sclerosis or limb-girdle muscular dystrophy.
ID5	No	Child with known trisomy 21, progressive gait disorder first noticed at the age of 7 years, increasingly limited walking distance since the age of 10 years (currently 500 m), confined to wheelchair for longer distances; CK ~ 1700 U/l, NLG unremarkable. Indication for diagnostic testing (neuromuscular panel): suspected myopathy
ID6	n.a.	abnormal ultrasound findings: increased nuchal translucency of 7 mm, vitium cordis (ductus venosus agenesis, unbalanced ventricles, mitral valve dysplasia, small apical ventricular septal defect, large echogenic focus in left ventricle, aberrant right subclavian artery and mild pericardial effusion); otherwise, timely fetal development; placental mosaic trisomy 18; retrospective findings in NGS data, diagnosis via newborn screening
ID7	No	Problems with stair climbing since the second year of life, more pronounced weakness from the 4th year of life with increased stumbling. Currently marked atrophy of thigh muscles, pseudohypertrophy of the calves. No fasciculation of tongue muscles. CK normal. Indication for diagnostic testing (exome): suspected limb girdle dystrophy or other neuromuscular disorder
ID8	No	First presentation in Germany at the age of 17. Hearing loss since birth, good communication with hearing aids. Anamnestically normal motor development in the first decade of life, proximal muscle atrophy and muscle weakness from the age of 10. Marked generalized muscular atrophy. Free walking still possible in the 17th year of life. Indication for diagnostic testing (exome): suspected limb girdle dystrophy
ID9	No "5q-SMA was preliminarily ruled out per MLPA"	Progressive muscle weakness predominantly affecting the lower limbs. Symptom onset with falls and difficulties climbing stairs at the age of 7, slowly progressive with worsening of gait, and also affected arm function, finally wheelchair-bound since the age of 35. Skoliosis. "5q-SMA preliminarily excluded per MLPA". Indication for diagnostic testing (neuromuscular panel): axonal neuropathy
ID10	No "5q-SMA was preliminarily ruled out per MLPA"	Onset of symptoms not exactly known (around 4 weeks of age), she was presented at 6 weeks of age with marked hypotonia and muscle weakness, rapid progressive course of disease, generalized hypotonia with severe muscle weakness, swallowing difficulties, respiratory failure, she died without treatment at 4 months: muscle biopsy: predominantly small immature muscle fibres, nearly all MHC/fetal/developmental positive, incipient fibre type grouping; "5q-SMA preliminarily ruled out per MLPA"

Abbreviations: ACMG = American College of Medical Genetics, CN = copy number, CNV = copy number variants, SNV = single nucleotide variants.

the context of panel or exome diagnostics, we established a 5q-SMA variant detection workflow based on *SMN2*-masked assembly in our bioinformatics pipeline and a specific *SMN1/2*-variant filtering to integrate it into our diagnostic set up.

We verified the accuracy to detect homozygous *SMN1* deletions of *SMN1* in a cohort of 212 patients with known SMA-MLPA results and the accuracy of variant calling in control samples with various CN of *SMN1* and *SMN2* in the masked SMN region

Table 2
Pathogenic *SMN1* variants detected in patient cohort

Patient	<i>SMN1</i> deletion	<i>SMN1</i> SNV	<i>SMN1</i> copies (by MLPA)	<i>SMN2</i> copies (by MLPA)	Sex	Current age	Age at onset	maximal motor achievement	current motor achievement
ID1	Heterozygous deletion	heterozygous c.821 C>T p.(Thr274Ile)	1	2	m	44 years	6 months	sitter	non-sitter
ID2	heterozygous deletion	heterozygous c.821 C>T p.(Thr274Ile)	1	2	m	18 years	2 years	walker	sitter
ID3	homozygous deletion		0	4	f	8 years	6 months	walker	walker
ID4	homozygous deletion		0	4	m	47 years	30 years	walker	sitter
ID5	homozygous deletion		0	4	m	12 years	7 years	walker	walker
ID6	homozygous deletion		0	1	n.d.	prenatal	n.a.	n.a.	n.a.
ID7	homozygous deletion		0	4	m	8 years	2 years	walker	walker
ID8	homozygous deletion		0	3	m	17 years	10 years	walker	walker
ID9		2 compound heterozygous SNVs c.469 C>T p.(Gln157*) and c.821 C>T p.(Thr274Ile)	1	1	f	48 years	7 years	walker	sitter
ID10		2 compound heterozygous SNVs c.549del p.(Lys184Serfs*29)and c.815A>G p.(Tyr272Cys)	2	2	f	death at 4.5 months of age	4 weeks	n.a.	n.a.

Abbreviations: MLPA = Multiplex Ligation-dependent Probe Amplification, n.a. = non applicable.

(Supplemental Table 1 and 2). Our results show that the new 5q-SMA variant detection workflow allows robust detection of a homozygous *SMN1* deletion. Of note, the workflow is not designed to estimate the absolute *SMN1* and *SMN2* CNs. Samples with *SMN1* CN ≥ 2 and samples with *SMN2* CN > 1 showed NM_000344.4:c.840 C>T VAFs within a similar overlapping range (Suppl. Table 1 and 2). Therefore individuals with only one copy of *SMN1* and a variable number of *SMN2* copies (CN ≥ 2) cannot be differentiated from individuals with two copies of *SMN1* and a variable number (≥ 2) of *SMN2* copies (≥ 2).

The application of the new 5q-SMA variant detection workflow in clinical cohorts allowed the identification of five patients who were not clinically suspected to have SMA when the diagnostic testing was initiated (Table 1, patients ID3, ID4, ID5, ID7, ID8). Consistent with a manifestation beyond infancy, all three patients carried four *SMN2* copies each, as determined by supplemental MLPA analysis. In one of the patients (ID4), the first symptoms uncharacteristically occurred after the age of 40 years. In other patients, the diagnosis may have been complicated by the pre-existing condition of trisomy 21 (ID5), or by concomitant hearing loss and a first clinical presentation late in the disease course (ID8). In one patient (ID6), a homozygous *SMN1* deletion could be shown after retrospective analysis of NGS data from prenatal trio exome sequencing. The prenatal exome had been initiated due to abnormal ultrasound findings with neck edema and a complex cardiac defect. Prenatally, there had been no evidence of neuromuscular symptoms in the unborn child. However, compatible with the presence of only one *SMN2* copy, the patient showed clinical symptoms of SMA at birth. In addition, an abnormal newborn screening result indicated the presence of a homozygous deletion of *SMN1*. The complex fetal cardiac defect might have been related to the diagnosis of SMA. It has been discussed in the literature that infantile SMA may be linked to congenital heart defects, especially in the presence of only one *SMN2* copy as in our patient [20]. Cardiac developmental defects may be a presymptomatic component to disease pathogenesis [21]. Due to the great therapeutic relevance of a prenatal SMA diagnosis, this case prompted us to test for the presence of a homozygous *SMN1* deletion in prenatal exome analyses if appropriate consent was obtained. However, no additional case was found in further 260 prenatal analyses. Even if prenatal SMA

is rare, screening for the presence of a homozygous *SMN1* deletion might nevertheless become routine part of prenatal exome analyses, as early diagnosis of this disease is essential for successful treatment.

The detection of SNVs within the *SMN1/SMN2* locus was also possible via our 5q-SMA variant detection workflow with *SMN2*-masked regions. In two patients, in whom the presence of a heterozygous *SMN1* deletion was known from a previous MLPA examination, a heteroallelic pathogenic SNV was detected. Given the suggestive clinical picture in these patients, the clinical diagnosis of 5q-SMA is likely. However, to definitely proof the diagnosis and to confirm that both variants (ACMG class 5) affect *SMN1* and not *SMN2* additional methods (specific long-range PCR or long-read sequencing) are required.

Of particular interest, two patients were identified who each carried two heterozygous pathogenic SNVs (ID9 and ID10). In both patients, the genetic diagnosis was made after a diagnostic odyssey: In the 48-year-old female patient (ID9), analysis of an axonal neuropathy panel had been initiated as the initial clinical diagnosis of 5q-SMA was not confirmed by MLPA which indicated absence of a hetero- or homozygous *SMN1* deletion. The second patient (ID10) was a four-month-old infant who showed generalized marked muscle weakness as well as respiratory insufficiency. Despite urgent clinical suspicion of 5q-SMA, a twofold testing for *SMN1* deletion had been negative. The subsequent muscle biopsy revealed predominantly small, immature muscle fibres (positive MHC fetal/developmental) and incipient fibre type grouping, so that a gene-panel for early infantile motor neuropathies was initiated that finally led to the diagnosis of 5q-SMA based on our 5q-SMA variant detection workflow. An analysis of parental samples confirmed that both SNVs affect different parental alleles in patient ID10, an analysis of parental samples was not possible in patient ID9. In the adult patient (ID9), we confirmed that the identified pathogenic SNVs are heteroallelic and localized in *SMN1* and not *SMN2* by long-range PCR [14] and subsequent long-read sequencing technology [15]. For the infant patient (ID10), high molecular weight DNA for long-read sequencing was not available. However, parental samples were available for long-read nanopore genomic sequencing to confirm that the identified SNVs were indeed associated with *SMN1* and were compound heterozygous in the index patient.

These patients illustrate that our 5q-SMA variant detection workflow detects SNVs but does not allow the phasing of variants. To prove that a variant is truly functionally significant as it affects *SMN1* and not its copy gene *SMN2*, confirmatory methods are necessary. Standard confirmatory methods such as long-range PCR of genomic *SMN1* followed by nested PCR and Sanger sequencing of each exon are laborious [14, 22]. Therefore, the lack of direct assignment of variants to *SMN1* or *SMN2* is a limitation of the presented workflow based on short-read sequencing. Ongoing advances in increasingly accurate long-read sequencing and bioinformatics may open new applications in clinical genome sequencing, not only resolving dark regions but also enabling direct phasing for haploptype identification [23–25].

Also in the presence of two pathogenic SNVs, there is an indication for therapy depending on the number of *SMN2* copies. In both patients with biallelic SNVs, one of their two variants suggests a complete allelic loss as a nonsense (patient ID7) or as a frameshift variant (patient ID8). *In trans* to the putative loss-of-function variant, both patients carried a missense variant. The patient with early infantile onset of symptoms (ID8) had two *SMN2* copies. The adult patient (ID7) had only one *SMN2* copy, which in combination with the relatively late manifestation of the disease may indicate a residual function of one of the alleles carrying the missense variant or an additional molecular rescue/modifier effect within or aside from *SMN2*. The correlation between the number of *SMN2* copies and the SMA phenotype severity is not always consistent in all patients described so far as a low number of *SMN2* copies in combination with a milder clinical phenotype has already been reported [26].

To our knowledge, no SMA patient carrying compound heterozygous sequence variants of the *SMN1* gene has been described in the literature to date. This could be due to the fact that sequencing of the *SMN1* gene is not possible with standard methods because of its high homology to the *SMN2* gene and, consequently, is probably often not performed when the homo- or heterozygous presence of an *SMN1* deletion has been excluded. The current German guidelines for 5q-SMA recommend sequencing of the *SMN1* gene in this situation only for patients with known parental consanguinity. In both patients, no consanguineous parents were known, consistent with the presence of compound heterozygous variants. Based on our results, biallelic variants in *SMN1* might be less rare than suggested by the current literature. As

consequence, the recommended diagnostic algorithm for SMA should be adapted to avoid overlooking 5q-SMA. In particular, clinicians and geneticists must be aware that a negative new-born screening for 5q-SMA does not rule out the diagnosis.

In conclusion, the here presented 5q-SMA variant detection workflow adapted for our NGS-based SMA diagnostic work up, is a reliable approach for the detection of homozygous deletions as well as heterozygous or homozygous SNVs within the challenging SMN locus on chromosome 5q13.2. The only limitations lie in the detection of healthy *SMN1* deletion carriers and the verification of *SMN2* copies, which is essential for timely SMA therapy and which still requires additional analysis by conventional MLPA.

ACKNOWLEDGMENTS

HE is grateful to the Faculty of Medicine of the Ludwig-Maximilians-Universität München for a fellowship.

AUTHOR CONTRIBUTION

Conception and Design: SK, VS, TW, AB-P, FS, TN, AA.

Administrative support: All authors.

Collection and assembly of data: All authors.

Data analysis and interpretation: SK, VS, TW, AB-P, FS, TN, AA.

Manuscript writing: SK, VS, TN, AA.

Final approval of manuscript: All authors.

FUNDING

This research was funded by the Medical Genetic Center without grants from government, commercial, or not-for-profit entities.

DATA AVAILABILITY STATEMENT

Full patients' data are not available publicly to respect participant privacy and consent. Anonymized data not published within this article will be made available by request from any qualified investigator.

CONFLICT OF INTEREST STATEMENT

All authors state no conflicts of interest to the content of this manuscript.

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: <https://dx.doi.org/10.3233/JND-221668>.

REFERENCES

- [1] Burr P, Reddivari AKR. Spinal Muscle Atrophy. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 [cited 2022 Nov 20]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK560687/>
- [2] Vill K, Schwartz O, Blaschek A, Gläser D, Nennstiel U, Wirth B, et al. Newborn screening for spinal muscular atrophy in Germany: Clinical results after 2 years. *Orphanet J Rare Dis.* 2021;16(1):153.
- [3] Wirth B, Karakaya M, Kye MJ, Mendoza-Ferreira N. Twenty-Five Years of Spinal Muscular Atrophy Research: From Phenotype to Genotype to Therapy, and What Comes Next. *Annu Rev Genomics Hum Genet.* 2020;21:231-61.
- [4] Munsat TL, Davies KE. International SMA consortium meeting. (26-28 June 1992, Bonn, Germany). *Neuromuscul Disord NMD.* 1992;2(5-6):423-8.
- [5] Mercuri E, Finkel RS, Muntoni F, Wirth B, Montes J, Main M, et al. Diagnosis and management of spinal muscular atrophy: Part 1: Recommendations for diagnosis, rehabilitation, orthopedic and nutritional care. *Neuromuscul Disord NMD.* 2018;28(2):103-15.
- [6] Finkel RS, Mercuri E, Meyer OH, Simonds AK, Schroth MK, Graham RJ, et al. Diagnosis and management of spinal muscular atrophy: Part 2: Pulmonary and acute care; medications, supplements and immunizations; other organ systems; and ethics. *Neuromuscul Disord NMD.* 2018;28(3):197-207.
- [7] Wirth B. An update of the mutation spectrum of the survival motor neuron gene (SMN1) in autosomal recessive spinal muscular atrophy (SMA). *Hum Mutat.* 2000;15(3):228-37.
- [8] Ebbert MTW, Jensen TD, Jansen-West K, Sens JP, Reddy JS, Ridge PG, et al. Systematic analysis of dark and camouflaged genes reveals disease-relevant genes hiding in plain sight. *Genome Biol.* 2019;20(1):97.
- [9] Mandelker D, Schmidt RJ, Ankala A, McDonald Gibson K, Bowser M, Sharma H, et al. Navigating highly homologous genes in a molecular diagnostic setting: A resource for clinical next-generation sequencing. *Genet Med Off J Am Coll Med Genet.* 2016;18(12):1282-9.
- [10] Köhler S, Gargano M, Matentzoglou N, Carmody LC, Lewis-Smith D, Vasilevsky NA, et al. The Human Phenotype Ontology in 2021. *Nucleic Acids Res.* 2021;49(D1):D1207-17.
- [11] 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 Off J Am Coll Med Genet.* 2015;17(5):405-24.
- [12] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma Oxf Engl.* 2009;25(14):1754-60.
- [13] Lai Z, Markovets A, Ahdesmaki M, Chapman B, Hofmann O, McEwen R, et al. VarDict: A novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res.* 2016;44(11):e108.
- [14] Kubo Y, Nishio H, Saito K. A new method for SMN1 and hybrid SMN gene analysis in spinal muscular atrophy using long-range PCR followed by sequencing. *J Hum Genet.* 2015;60(5):233-9.
- [15] Dai M, Xu Y, Sun Y, Xiao B, Ying X, Liu Y, et al. Revealing diverse alternative splicing variants of the highly homologous SMN1 and SMN2 genes by targeted long-read sequencing. *Mol Genet Genomics MGG.* 2022;297(4):1039-48.
- [16] Tham CY, Tirado-Magallanes R, Goh Y, Fullwood MJ, Koh BTH, Wang W, et al. NanoVar: Accurate characterization of patients' genomic structural variants using low-depth nanopore sequencing. *Genome Biol.* 2020;21(1):56.
- [17] Li H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinforma Oxf Engl.* 2018;34(18):3094-100.
- [18] Chen Y, Ye W, Zhang Y, Xu Y. High speed BLASTN: An accelerated MegaBLAST search tool. *Nucleic Acids Res.* 2015;43(16):7762-8.
- [19] Dangouloff T, Vrščaj E, Servais L, Osredkar D, SMA NBS World Study Group. Newborn screening programs for spinal muscular atrophy worldwide: Where we stand and where to go. *Neuromuscul Disord NMD.* 2021;31(6):574-82.
- [20] Rudnik-Schöneborn S, Heller R, Berg C, Betzler C, Grimm T, Eggermann T, et al. Congenital heart disease is a feature of severe infantile spinal muscular atrophy. *J Med Genet.* 2008;45(10):635-8.
- [21] Motyl AAL, Faller KME, Groen EJM, Kline RA, Eaton SL, Ledahawsky LM, et al. Pre-natal manifestation of systemic developmental abnormalities in spinal muscular atrophy. *Hum Mol Genet.* 2020;29(16):2674-83.
- [22] Eggermann K, Gläser D, Abicht A, Wirth B. Spinal muscular atrophy (5qSMA): Best practice of diagnostics, newborn screening and therapy. *Med Genet.* 2020;32(3):263-72.
- [23] Wagner J, Olson ND, Harris L, McDaniel J, Cheng H, Fungtammasan A, et al. Curated variation benchmarks for challenging medically relevant autosomal genes. *Nat Biotechnol.* 2022;40(5):672-80.
- [24] Chen X, Harting J, Farrow E, Thiffault I, Kasperaviciute D, Genomics England Research Consortium, et al. Comprehensive SMN1 and SMN2 profiling for spinal muscular atrophy analysis using long-read PacBio HiFi sequencing. *Am J Hum Genet.* 2023;110(2):240-50.
- [25] Edwards C, Martin B, Roberts M, Kempainen J, Routsong R, Killinger B, et al. P455: Long-range PCR and nanopore sequencing method resolves F8, GBA, CYP21A2, SMN1, and TNXB variants using a single streamlined workflow. *Genet Med Open.* 2023;1(1):100502.
- [26] Ludolph AC, Wurster CD. Therapeutic advances in SMA. *Curr Opin Neurol.* 2019;32(5):777-81.
- [27] Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A, Mesirov JP. Variant Review with the Integrative Genomics Viewer. *Cancer Res.* 2017;77(21):e31-4.