# scientific reports

OPEN

# Comparative study of third-generation sequencing-based CASMA-trio and STR linkage analysis for identifying *SMN1* 2 + 0 carriers

Jianchun He[1,2,3,4,6], Wenzhi He[1,2,3,4,6], Jiajia Xian[1,2,3,4], Yanchao Wang[1,2,3,4], Xiaoyan Ma[1,2,3,4], Shuangru Lin[1,2,3,4], Jianfan Chen[1,2,3,4], Gaoyuan Xu[1,2,3,4], Qingqing Chen[2,3,4,5], Qing Li[1,2,3,4✉] & Shaoying Li[1,2,3,4✉]

Spinal Muscular Atrophy (SMA, MIM#253300) is an autosomal recessive neuromuscular disorder caused by defects in the Survival Motor Neuron (*SMN*) gene. The *SMN1* gene, recognized as the primary pathogenic gene for SMA, exhibits a high degree of sequence homology with *SMN2* gene. Individuals with the *SMN1* 2 + 0 genotype represent a unique type of SMA carrier, characterized by two SMN1 copies on one chromosome and zero copies on the other. Accurate identification of this type of carrier is crucial for genetic counseling in families. This study included 28 samples from five SMA families, each with an affected patient carrying a homozygous deletion of the SMN1 gene and a parent suspected to be a *SMN1* 2 + 0 carrier. Comprehensive Analysis of SMA (CASMA), based on third-generation sequencing technology, was used to detect the *SMN1* and *SMN2* copy numbers in the samples, and *SMN1* 2 + 0 carriers were identified through *SMN1* haplotypes in parent-child trios (CASMA-trio). The results were compared with those obtained using Multiplex Ligation-dependent Probe Amplification (MLPA) combined with Short Tandem Repeat (STR) linkage analysis. The *SMN1* and *SMN2* copy numbers detected by MLPA and CASMA were concordant across 25 peripheral blood samples, whereas CASMA failed to accurately determine the copy numbers in the remaining 3 amniotic fluid samples. CASMA-trio identified 5 members from 4 families as *SMN1* 2 + 0 carriers, which were consistent with the results from STR linkage analysis. However, the two methods yielded inconsistent results for the proband's father in one family. These findings suggest that CASMA has the potential to detect *SMN1* and *SMN2* copy numbers. Compared to STR linkage analysis, CASMA-trio only requires a parent-child trio to analyze *SMN1* 2 + 0 carriers, demonstrating a broader application prospect. Implementing CASMA-trio can facilitate comprehensive screening for SMA carriers.

**Keywords** Spinal muscular atrophy, *SMN1* 2 + 0 carrier, STR linkage analysis, Third-generation sequencing

Spinal Muscular Atrophy (SMA, MIM#253300) is a common autosomal recessive neuromuscular disorder caused by the degeneration of spinal anterior horn motor neurons, primarily due to the deletion and mutation of the Survival Motor Neuron (*SMN*) gene[1]. The clinical features of SMA include the degeneration and loss of lower motor neurons in the spinal cord and brainstem nuclei, leading to progressive symmetrical proximal muscle weakness and atrophy of the lower limbs, which can affect multiple systems[2]. The incidence of SMA is approximately 1/6,000 to 1/10,000, making it one of the most common genetic causes of childhood mortality[3].

[1]Department of Obstetrics and Gynecology, Experimental Department of Obstetrics and Gynecology Institute, The Third Affiliated Hospital of Guangzhou Medical University, Guangzhou 510150, Guangdong, China. [2]Guangdong Provincial Key Laboratory for Major Obstetric Diseases, Guangzhou 510150, Guangdong, China. [3]Guangdong Provincial Clinical Research Center for Obstetrics and Gynecology, Guangzhou 510150, Guangdong, China. [4]Guangdong-Hong Kong-Macao Greater Bay Area Higher Education Joint Laboratory for Maternal-Fetal Medicine, Guangzhou 510150, Guangdong, China. [5]Department of Obstetrics and Gynecology, Department of Obstetrics, The Third Affiliated Hospital of Guangzhou Medical University, Guangzhou 510150, Guangdong, China. [6]These authors have contributed equally to this work. ✉email: 81292522@163.com; amyleewang@163.com

Type I SMA is the most severe form, presenting with significant muscle weakness before six months of age, with a median survival time of 24 months[4,5]. Although medications exist that can halt the progression of SMA and significantly improve motor function in patients, these drugs are expensive and have limited availability in most regions globally[6,7]. Therefore, the American College of Medical Genetics and Genomics (ACMG) recommends population-based carrier screening[8]. If both partners are identified as SMA carriers, subsequent prenatal diagnosis during pregnancy can prevent the birth of affected children, significantly reducing the incidence of the disease.

The *SMN* genes include *SMN1* and *SMN2*, which are highly homologous. *SMN1* and *SMN2* are located in two duplicated 500-kb segments on the 5q13 chromosomal locus and are prone to rearrangements and deletions[9,10]. *SMN1* is the primary causative gene for SMA, with approximately 95% of patients resulting from homozygous deletions of *SMN1*[11,12]. Normal individuals have two copies of the *SMN1* gene, one on each chromosome, with a genotype of 1 + 1; SMA patients have 0–1 copies, with genotypes of 0 + 0 or 0 + 1^D (1^D indicates a small pathogenic variant on one chromosome). The majority of SMA carriers have a 1 + 0 genotype, while a small proportion have 1 + 1^D and 2 + 1^D genotypes. There is a special type of SMA carrier who possesses two *SMN1* copies on one chromosome and zero copies on the other, known as a 2 + 0 carrier. Reports indicate that 2 + 0 carriers occur in approximately 4% of the SMA carrier population[10,13]. Therefore, effectively identifying 2 + 0 carriers is important for increasing the detection rate of SMA carriers and assessing the risk of SMA in the next generation.

Currently, molecular biology techniques such as quantitative PCR, Multiplex Ligation-dependent Probe Amplification (MLPA)[14,15], digital PCR, TaqMan quantitative technology, and next-generation sequencing[16–18] are used to screen for *SMN1* 1 + 0 carriers. Clinically, MLPA is the gold standard for SMA genetic testing. Although MLPA has many advantages, it cannot detect point mutations or special *SMN1* 2 + 0 carriers. Typically, *SMN1* 2 + 0 carriers are inferred in families with one affected child and one parent with two *SMN1* copies[19,20]. However, this method relies on the presence of a SMA proband, making it impractical for carrier screening. Another alternative method is Short Tandem Repeat (STR) linkage analysis, which identifies *SMN1* 2 + 0 carriers through haplotype analysis of three-generation families or multi-child families[21]. However, the clinical feasibility of this method is limited due to the need for multiple family members and the complexity of STR analysis.

A recently developed method, Comprehensive Analysis of SMA (CASMA), based on long-range PCR and third-generation sequencing, analyzes the complete *SMN1/2* sequence and can simultaneously screen for *SMN1* 1 + 0, 1 + 1^D, and 2 + 1^D carriers[22]. CASMA can also obtain the full-length haplotype of *SMN1* and, combined with trio family analysis, effectively deduce 2 + 0 carriers. This method, known as CASMA-trio, does not rely on the presence of a proband, nor does it require SNP analysis of three generations or specific population SNP haplotype construction, making it an effective and broadly applicable method for detecting *SMN1* 2 + 0 carriers[23].

This study includes five SMA families, where one parent of the proband is suspected to be a 2 + 0 carrier. Through *SMN* gene copy number analysis combined with STR linkage analysis and CASMA-trio, we aimed to determine the genotype of suspected 2 + 0 carriers, validate the effectiveness of CASMA-trio in analyzing 2 + 0 carriers, and explore the clinical application value of CASMA-trio in SMA carrier screening.

## Materials and methods
### Subjects
The study included a total of 28 samples from 5 families, including three fetal amniotic fluid samples. The copy number of *SMN1* and *SMN2* were analyzed using SALSA MLPA Probemix P060-B2 (MRC-Holland) and Coffalyser software. All enrolled families met the following criteria: (1) one SMA patient was identified within the family; (2) Each family included at least one trio of father, mother, and child. The study was approved by the ethics committee of the Third Affiliated Hospital of Guangzhou Medical University (Ethics Approval [2021] No. 254), and written informed consent was signed by all participating family members.

### Fluorescent PCR-Capillary electrophoresis (PCR/CE)
The multiple PCR capillary electrophoresis DNA fragment analysis technique was used to simultaneously amplify 14 genetic markers in one PCR reaction, with one primer in each marker pair labeled with FAM fluorophore. PCR products were separated by capillary electrophoresis to obtain peak heights and genotypes of polymorphic markers. The 14 genetic markers include 4 markers for *SMN* copy number testing (1 marker for determining the total copy number of *SMN1* and *SMN2*, 1 marker for assessing the copy number of exon 7 in *SMN1* and *SMN2* separately, 1 marker for determining the total copy number of exon 8 in *SMN1* and *SMN2*, 1 marker is a single-copy reference gene), 2 markers for experimental quality control, and 8 closely linked molecular tags for *SMN1* and *SMN2*, labeled S01 to S08. S04, S06, and S07 are molecular tags tightly linked to *SMN2* located within 1 Mb upstream the gene, and S01, S02, S03, S05, and S08 are molecular tags tightly linked to *SMN1* located within 1 Mb downstream the gene. Their relative positions are shown in Supplementary Fig.S1 online and are used for linkage analysis of SMA families. Operations were performed according to the manufacturer's instructions for the motor neuron survival gene copy number detection kit (fluorescent PCR-capillary electrophoresis method) from Jingzhun Company.

### Full-length sequencing of SMN1/2
The full-length sequencing of *SMN1* and *SMN2* genes was conducted using the CASMA method, with long-range PCR, library preparation, and long-read sequencing as described by Li S[22]. The full length of *SMN1* and *SMN2* genes (*SMN1/2*-FL, 28.5 kb) were amplified from genomic DNA samples, Subsequently, a one-step end repair and ligation protocol was employed to attach unique PacBio barcoded adapters to PCR products, followed by digestion with exonucleases to remove failed ligation DNA. Each of the above reaction mixtures from a single sample was purified, quantified, and then pooled with equal mass to form pre-library. The sequencing primers are annealed and the polymerase added to the pre-library using the Sequel II Binding Kit v.3.2 (Pacific

Biosciences) to prepare single-molecule real-time dumbbell (SMRTbell) library, followed by 30-hour sequencing on the Sequel IIe platform (Pacific Biosciences) in circular consensus sequencing (CCS) mode.

### PacBio data analysis for SMN1/2 haplotype and copy number

The raw sequencing subreads were processed to obtain high-fidelity CCS reads, decoded into individual samples using unique barcodes, and aligned to the hg38 reference genome using the software suite (smrtlink 10.1.0.119588, Pacific Biosciences). The *SMN1* and *SMN2* genes were distinguished by the c.840 functional homologous sequence variant. For haplotype analysis of *SMN1* and *SMN2*, each CCS read was aligned to the hg38 reference genome to detect any SNPs. SNPs present in repeat regions were masked. SNPs with variant allele frequencies < 20% or > 80% were filtered out to generate a SNP matrix on each CCS read. The CCS reads were recursively divided into two groups using SNPs until no further grouping was possible. Each final group represented a specific haplotype, and the read count for each haplotype was obtained. Haplotype assignments for all samples were also confirmed by visual inspection of the CCS reads in the Integrative Genomics Viewer (IGV). The copy number of *SMN1*/2 was calculated using a Poisson distribution-based caller, based on the number and read counts of each haplotype, as reported by Li S[22].

The full-length sequencing of *SMN1/2* and PacBio data analysis were performed by Berry Genomics Corporation.

### SMN1 2 + 0 analysis by CASMA-trio[23]

The *SMN1* 2 + 0 analysis requires a family trio consisting of parents and their child, utilizing the *SMN1*-FL haplotype in the analysis process. Individuals with two copies of *SMN1* are used for the 2 + 0 analysis. If the subject being analyzed is the child, there are two possible scenarios: (1) the child's *SMN1* haplotypes are inherited from only one parent, indicating that the child is an *SMN1* 2 + 0 carrier; (2) the child's *SMN1* haplotypes are inherited from both parents, indicating that the child is not an *SMN1* 2 + 0 carrier. If the subject being analyzed is a parent, there are three possible scenarios: (1) all or (2) none of the parent's *SMN1* haplotypes are inherited by the child, indicating that the parent is an *SMN1* 2 + 0 carrier; (3) part of the parent's *SMN1* haplotypes are inherited by the child, indicating that the parent is not an *SMN1* 2 + 0 carrier.

## Results
### Detection of SMN1 and SMN2 copy numbers

To assess the performance of CASMA in detecting *SMN1* and *SMN2* copy numbers, both MLPA and CASMA were applied concurrently to analyze the 28 samples. However, three amniotic fluid samples (1110D, 479 A, and KJS41) could not be effectively assessed by CASMA. For the remaining 25 peripheral blood samples, the copy numbers of *SMN1* and *SMN2* detected by CASMA were consistent with those obtained by MLPA. (Table 1).

### Analysis of SMN1 2 + 0 carriers by STR Linkage Analysis and CASMA-trio

To determine whether subjects with two copies of *SMN1* were 2 + 0 carriers, STR linkage analysis and CASMA-trio analysis were performed on their families (Table 1 and Supplementary Table 1).

In Family 1, the couple had a fetus (III-1) affected by SMA. MLPA analysis showed that the father (II-1) and maternal grandfather (I-1) each had one copy of *SMN1*, the mother (II-2) had two copies of *SMN1*, and the maternal grandmother (I-2) had three copies of *SMN1*. STR linkage analysis suggested that the mother was an *SMN1* 2 + 0 carrier (Fig. 1a). To confirm this, the mother underwent CASMA-trio analysis, examining the mother, maternal grandfather, and maternal grandmother. The results showed that both of the mother's *SMN1* haplotypes were inherited from the maternal grandmother, confirming her as an *SMN1* 2 + 0 carrier (Fig. 1c), consistent with the STR linkage analysis of the three-generation family with the proband.

In Family 2, similar to Family 1, STR linkage analysis combined with CASMA-trio confirmed that the mother (II-2) was an *SMN1* 2 + 0 carrier, with both *SMN1* haplotypes inherited from the maternal grandfather (I-1) (Fig. 1b and d).

In Family 3, MLPA analysis indicated that the father, son, and maternal grandfather were *SMN1* 1 + 0 carriers, while the mother had two copies of *SMN1* and the maternal grandmother had three copies of *SMN1*. Due to the lack of a proband sample, it was not possible to definitively determine whether the mother was an *SMN1* 2 + 0 carrier, necessitating the analysis of more family members. The sister 1 (II-1) and sister 2 (II-2) had one and three copies of *SMN1*, respectively. STR linkage analysis showed that the mother and sister 1 inherited the father's 0 copy *SMN1* gene, while the mother and sister 2 inherited the 2 copy *SMN1* gene from the maternal grandmother, confirming the mother as an *SMN1* 2 + 0 carrier (Fig. 2a). CASMA-trio analysis of the mother, maternal grandfather, and maternal grandmother confirmed the mother as an *SMN1* 2 + 0 carrier and the sister 3 (II-3) as a 1 + 1 normal individual (Fig. 2b).

In Family 4, the couple had a fetus affected by SMA (II-1) and a healthy daughter (II-2). MLPA analysis indicated that the daughter had two copies of *SMN1*, similar to the mother (I-2), while the father (I-1) was an *SMN1* 1 + 0 carrier. Due to the unavailability of samples from the maternal grandparents, three-generation linkage analysis could not determine the mother's genotype. STR linkage analysis indicated that both the fetus and the daughter inherited the father's 0-copy *SMN1* gene, suggesting that the daughter's two-copy *SMN1* gene was inherited from the mother, implying the mother was an *SMN1* 2 + 0 carrier, and the daughter was also an *SMN1* 2 + 0 carrier (Fig. 3a). CASMA-trio results showed that the daughter inherited all haplotypes from the mother, confirming both the mother and the daughter as *SMN1* 2 + 0 carriers (Fig. 3b).

Notably, in Family 5, the couple had a son affected by SMA (III-1). Amniotic fluid MLPA results showed that the fetus (III-2) had one copy of *SMN1*, and the mother (II-2) was an *SMN1* 1 + 0 carrier, while the father (II-1) had two copies of *SMN1*, raising suspicion that the father might be an *SMN1* 2 + 0 carrier. Further testing of the paternal grandparents (I-1 and I-2) revealed that both had two copies of *SMN1*. Combining MLPA results and

| Family | Member | Sample | MLPA SMN1 copy | MLPA SMN2 copy | CASMA SMN1 copy | CASMA SMN2 copy | PCR/CE SMN1 pattern | CASMA-trio SMN1 pattern |
|---|---|---|---|---|---|---|---|---|
| 1 | I–1(Maternal grandfather) | KJS33 | 1 | 2 | 1 | 2 | 1 + 0 | 1 + 0 |
| | I–2(Maternal grandmother) | KJS32 | 3 | 0 | 3 | 0 | 2 + 1 | 2 + 1 |
| | II–1(Father) | KJS31 | 1 | 2 | 1 | 2 | 1 + 0 | 1 + 0 |
| | II–2(Mother) | KJS30 | 2 | 1 | 2 | 1 | 2 + 0 | 2 + 0 |
| | III–1(Fetus(affected)) | 1110D | 0 | 2 | N/A | N/A | 0 + 0 | N/A |
| 2 | I–1(Maternal grandfather) | KJS70 | 3 | 1 | 3 | 1 | 2 + 1 | 2 + 1 |
| | I–2(Maternal grandmother) | KJS69 | 1 | 3 | 1 | 3 | 1 + 0 | 1 + 0 |
| | II–1(Father) | KJS67 | 1 | 2 | 1 | 2 | 1 + 0 | 1 + 0 |
| | II–2(Mother) | KJS66 | 2 | 2 | 2 | 2 | 2 + 0 | 2 + 0 |
| | III–1(Daughter(affected)) | KJS68 | 0 | 3 | 0 | 3 | 0 + 0 | 0 + 0 |
| 3 | I–1(Maternal grandfather) | KJS51 | 1 | 1 | 1 | 1 | 1 + 0 | 1 + 0 |
| | I–2(Maternal grandmother) | KJS50 | 3 | 2 | 3 | 2 | 2 + 1 | 2 + 1 |
| | II–1(Sister1) | KJS55 | 1 | 1 | 1 | 1 | 1 + 0 | 1 + 0 |
| | II–2(Sister2) | KJS52 | 3 | 2 | 3 | 2 | 2 + 1 | 2 + 1 |
| | II–3(Sister3) | KJS53 | 2 | 2 | 2 | 2 | 1 + 1 | 1 + 1 |
| | II–4(Mother) | KJS48 | 2 | 1 | 2 | 1 | 2 + 0 | 2 + 0 |
| | II–5(Father) | KJS49 | 1 | 3 | 1 | 3 | 1 + 0 | 1 + 0 |
| | III–1(Son) | KJS54 | 1 | 1 | 1 | 1 | 1 + 0 | 1 + 0 |
| 4 | I–1(Father) | KJS43 | 1 | 1 | 1 | 1 | 1 + 0 | 1 + 0 |
| | I–2(Mother) | KJS44 | 2 | 2 | 2 | 2 | 2 + 0 | 2 + 0 |
| | II–1(Fetus(affected)) | 479 A | 0 | 2 | N/A | N/A | 0 + 0 | N/A |
| | II–2(Daughter) | KJS45 | 2 | 0 | 2 | 0 | 2 + 0 | 2 + 0 |
| 5 | I–1(Paternal grandfather) | KJS37 | 2 | 2 | 2 | 2 | 2 + 0 | 1 + 1 |
| | I–2(Paternal grandmother) | KJS38 | 2 | 2 | 2 | 2 | 2 + 0 | 1 + 1 |
| | II–1(Father) | KJS35 | 2 | 2 | 2 | 2 | 2 + 0 | 1 + 1 |
| | II–2(Mother) | KJS34 | 1 | 3 | 1 | 3 | 1 + 0 | 1 + 0 |
| | III–1(Son(affected)) | KJS36 | 0 | 3 | 0 | 3 | 0 + 0 | 0 + 0(de novo) |
| | III–2(Fetus) | KJS41 | 1 | 2 | N/A | N/A | 1 + 0 | N/A |

**Table 1**. Copy number of *SMN1*/2 and distribution pattern of *SMN1* in 28 samples from 5 families. N/A: not available.

STR linkage analysis, the father and paternal grandparents were likely *SMN1* 2 + 0 carriers. However, the *SMN1* 2 + 0 genotype in father cannot be directly identified due to the limited informative loci in STR linkage analysis and was speculated according to the genotypes of the proband and the mother (Fig. 4a). CASMA identified that the father harbored two copies *SMN1*. Trio analysis revealed that Allele 1 was inherited from paternal grandmother and Allele 2 was inherited from paternal grandfather. Therefore, the father has a normal *SMN1* 1 + 1 genotype. CASMA identified that the mother has an *SMN1* 1 + 0 genotype, and the proband has an *SMN1* 0 + 0 genotype (Fig. 4b). The *SMN1* null allele in fetus, detected by MLPA, can only be inherited from the mother.

## Discussion

For families with homozygous deletions in the *SMN1* gene, it is crucial to determine the genotypes of both parents to assess the recurrence risk of SMA when one parent is a classic 1 + 0 carrier and the other has two copies of the *SMN1* gene. However, genetic screening for *SMN1* 2 + 0 carriers, who make up about 4% of all SMA carriers, remains a global challenge.

In recent years, single-molecule long-read TGS technology has provided new solutions for detecting complex genetic diseases such as thalassemia, congenital adrenal hyperplasia, and fragile X syndrome[24–26]. TGS technology has also been applied to SMA gene analysis, leading to the development of the CASMA method. Studies have shown that CASMA has a 99.4% accuracy in determining *SMN1* and *SMN2* copy numbers in 337 clinical samples compared to MLPA, NGS, and qPCR tests[22]. CASMA accurately analyzes *SMN1* and *SMN2* copy numbers by sequencing the full-length *SMN1* and *SMN2* genes and combining haplotypes and read counts in a Poisson model. This study tested the *SMN1*/2 copy numbers in 28 samples, and the results from MLPA and CASMA were consistent in 25 peripheral blood samples, except for 3 amniotic fluid sample. The 1110D and 479 A amniotic fluid samples could not be effectively analyzed due to low DNA concentrations, which resulted in insufficient sample amounts to meet the testing requirements. The detection failure in the KJS41 amniotic fluid sample was likely attributed to poor DNA integrity extracted from amniotic fluid, thereby hindering the successful amplification of the full-length *SMN1* fragment via long-range PCR. These results underscore the inherent limitations of prenatal samples and highlight the technical challenges for the application of CASMA in
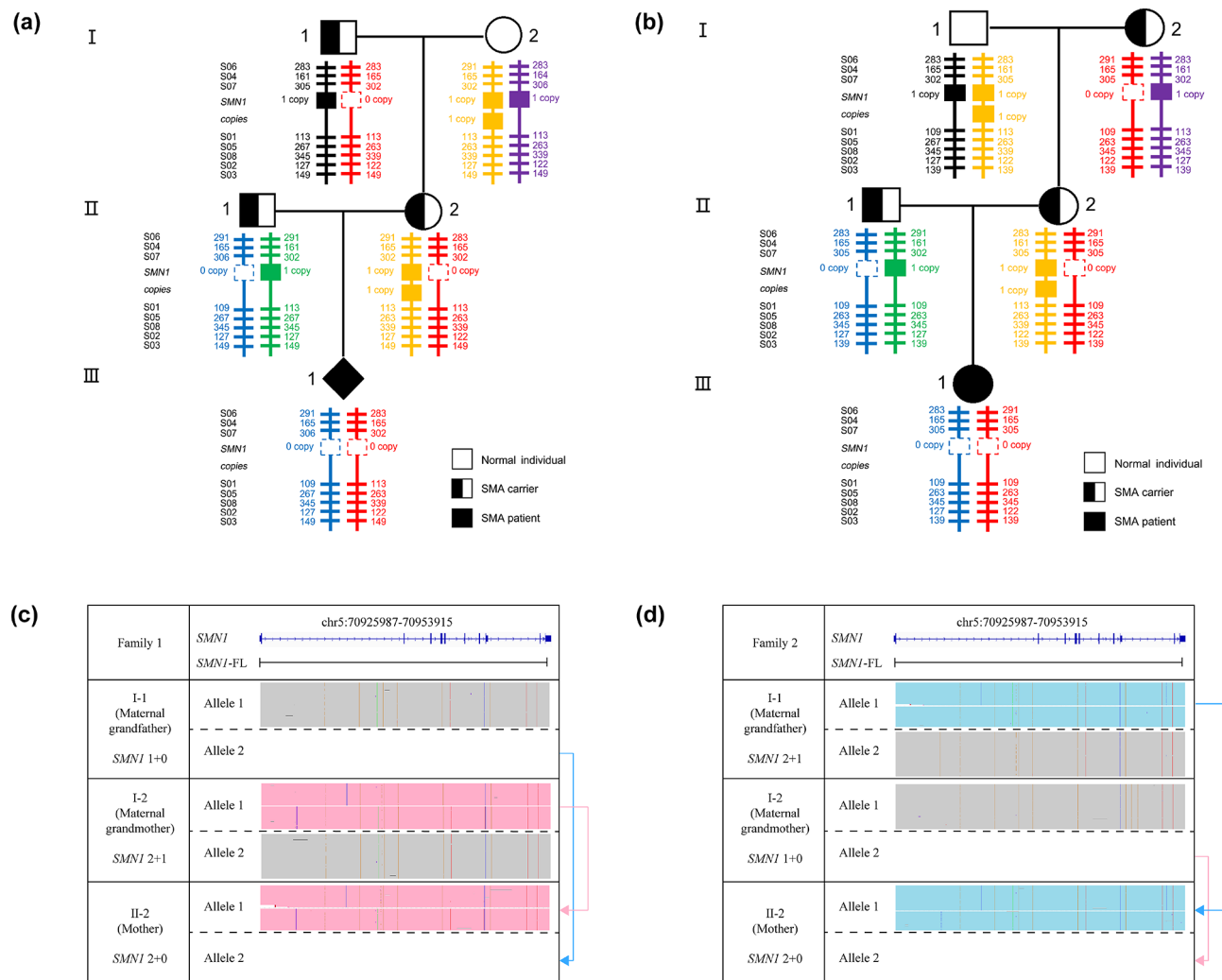
**Fig. 1**. STR linkage analysis and CASMA-trio analysis pattern graph of 2 + 0 carriers in three-generation family with a proband (**a**) (**b**) STR linkage analysis pattern graph of Family 1 and Family 2. The mother's genotype was inferred based on the proband and the genotypes of the maternal grandparents. Vertical lines of different colors represent chromosomes from different origins. The dashed boxes on the vertical lines indicate 0 copy, while the solid boxes represent 1 copy. S06, S04, S07, S01, S05, S08, S02, and S03 represent the eight molecular marker loci, and the numbers in the graph indicate the PCR amplification fragment lengths (bp) of each molecular marker locus; (**c**) IGV plot of Family 1, (**d**) IGV plot of Family 2. (**c**) and (**d**) show the SMN1 -FL haplotypes. Arrows indicate the inheritance of alleles across generations.

prenatal SMA diagnosis, suggesting potential directions for further optimization of the method. One potential direction for further optimization of amniotic fluid samples with insufficient DNA template or partially degraded DNA is to increase the input and number of amplification cycles to improve the yield of the amplicons, or to increase the pooling mass to get more eligible full-length reads. The other potential approach would be to amplify shorter amplicons including exons 7 and 8 of *SMN* genes, and analyze them together with the full-length *SMN* genes of parents to identify the fetal genotypes.

Although quantitative gene detection techniques such as MLPA and Real-time PCR can determine *SMN1* and *SMN2* copy numbers, they cannot distinguish between the *SMN1* 2 + 0 genotype and the normal *SMN1* 1 + 1 genotype, requiring additional detection techniques. As early as 1999, Chen et al. confirmed the existence of the *SMN1* 2 + 0 genotype using fluorescence quantitative analysis combined with haplotype analysis of the *SMN1* gene[27]. In 2001, Mailman et al. detected the *SMN1* 2 + 0 genotype using cell fusion technology, but this technique is time-consuming and labor-intensive[19]. Recently, single-sperm sequencing has been used to clarify the *SMN1* 2 + 0 genotype in males, but it is not applicable to females[28]. In this study, STR linkage analysis was performed for all 25 analyzed subjects. Among them, 9 individuals were identified as having two copies of *SMN1* and were speculated to be *SMN1* 2 + 0 carriers based on pedigree analysis, which was ultimately confirmed by CASMA-trio. In Family 5, STR linkage analysis based on the genotypes of the proband and the mother suggested that individuals I−1, I−2, and II−1 were *SMN1* 2 + 0 carriers. However, CASMA-trio confirmed that I−1, I−2 and II−1 were normal *SMN1* 1 + 1 genotypes, II−2 and III-2 has *SMN1* 1 + 0 genotypes, and III-1 had a *SMN1*
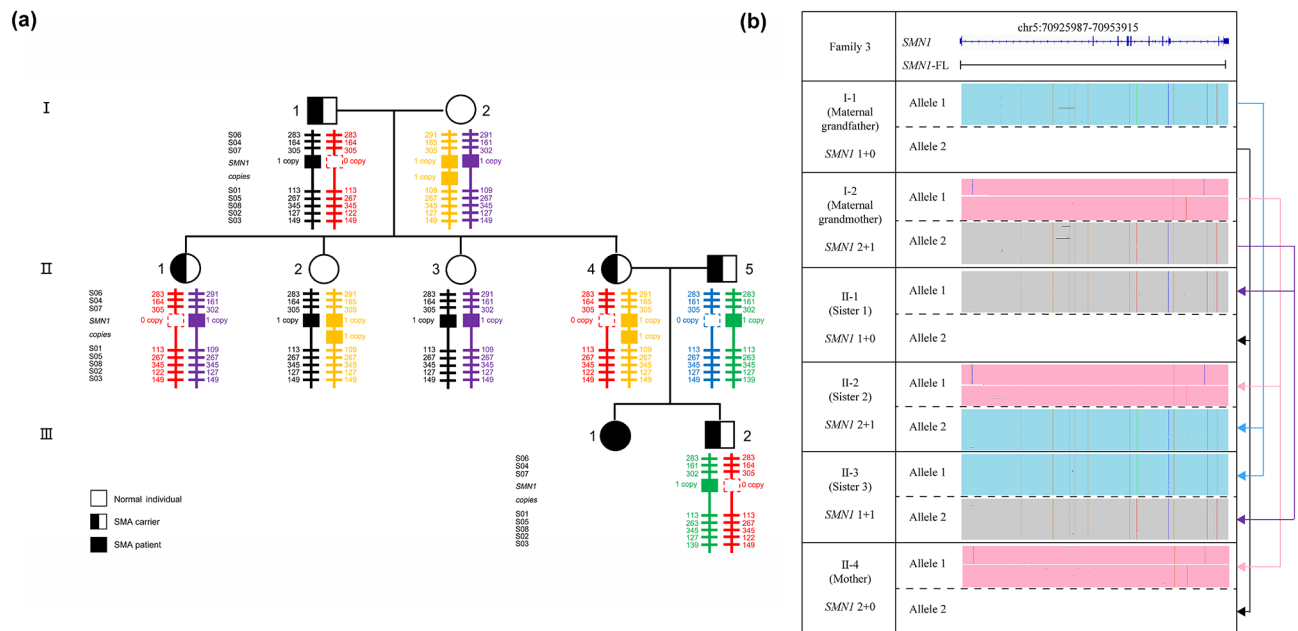
**Fig. 2**. STR linkage analysis and CASMA-trio analysis pattern graph of 2 + 0 carriers in three-generation family without a proband (**a**) STR linkage analysis pattern graph of Family 3. The mother's genotype was inferred based on the genotypes of the maternal grandparents, maternal sisters, and the son. Vertical lines of different colors represent chromosomes from different origins. The dashed boxes on the vertical lines indicate 0 copy, while the solid boxes represent 1 copy. S06, S04, S07, S01, S05, S08, S02, and S03 represent the eight molecular marker loci, and the numbers in the graph indicate the PCR amplification fragment lengths (bp) of each molecular marker locus; (**b**) IGV plot of Family 3, showing the SMN1-FL haplotypes, with arrows indicating the inheritance of alleles across generations.

0 + 0 genotype. This phenomenon could be due to extensive repetitive sequences surrounding the *SMN1* and *SMN2* loci, predisposing the region to unequal crossover or homologous recombination events, resulting in de novo deletions[29].

Based on the CASMA-trio results, we conducted a reanalysis of STR linkage in Family 5 and identified potential homologous recombination in the maternal haplotype of III-1 (Supplementary Fig. S2).

Due to the limited resolution of STR linkage markers, we further employed SNP array linkage analysis to determine the potential homologous recombination in affected son (Supplementaty Fig. S3). The results revealed homologous recombination events in both the paternal and maternal haplotypes of the affected son, with recombination sites spanning the *SMN1* gene region. These recombination events made it challenging to accurately determine the genotype of all the family members using conventional genetic methods. A comprehensive analysis was performed based on the results of CASMA-trio assay combined with the SNP array testing, revealing the following: (1) Potential homologous recombination between the maternal orange and green chromosomes may occur upstream of *SMN1*. In this context, the recombinant chromosome in the affected son lacks *SMN1*, suggesting that the non-recombinant orange chromosome is also null for *SMN1*. For the *SMN1* 1 + 0 carrier fetus (III-2), the zero-copy *SMN1* is located on the maternal orange chromosome, while the paternal red chromosome carries one copy of *SMN1*. Under these circumstances, the father (II-1) have a *SMN1* 1 + 1 genotype, which is consistent with the CASMA results. Conversely, if recombination between the maternal orange and green chromosomes occurred downstream of *SMN1*, the non-recombinant maternal green chromosome would lack *SMN1*. while the non-recombinant maternal orange chromosome would retain one copy of *SMN1*. In this context, the paternal red chromosome in the *SMN1* 1 + 0 carrier fetus (III-2) would also lack *SMN1*, as the maternal orange chromosome contributes one *SMN1* copy. Consequently, the father (II-1) would have a 2 + 0 *SMN1* genotype, a hypothesis refuted by CASMA-trio analysis though accurate sequencing, variant calling, and haplotype analysis. (2) Based on the current findings, it is difficult to determine whether potential homologous recombination between the paternal blue and red chromosomes occurs upstream or downstream of *SMN1*. Regardless of whether the paternal recombinant chromosome occur uptream or downsteam of *SMN1*, the genotypes of predigree were consistent with the results of CASMA-trio analysis. Determining the recombination region requires additional supportive data.

CASMA-trio achieved effective and universal screening for *SMN1* 2 + 0 carriers through family trio analysis. Studies have shown that CASMA-trio successfully identified whether subjects with two copies of *SMN1* ($n = 16$) and *SMN2* ($n = 43$) were in the 2 + 0 mode[23]. Importantly, CASMA-trio does not rely on the presence of an affected proband, population-specific SNPs or haplotypes, or complex SNP analysis involving three-generation family members. There are some limitations to CASMA-trio analysis. Firstly, the effectiveness of this method requires trio analysis, which is not always readily available. In cases where only one parent is available, a definitive
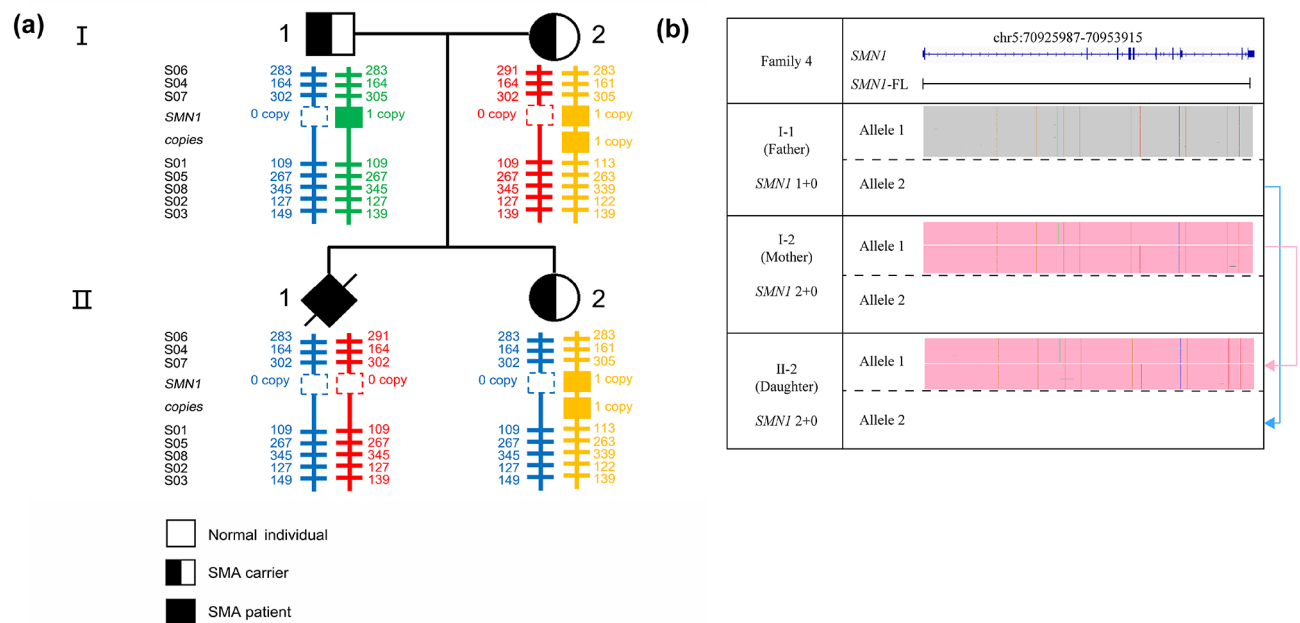
**Fig. 3**. STR linkage analysis and CASMA-trio analysis pattern graph of 2 + 0 carriers in two-generation family with a proband (**a**) STR linkage analysis pattern graph of Family 4. The mother's genotype was inferred through two-generation analysis using the proband and the healthy daughter with two copies of the SMN1 gene as references. Vertical lines of different colors represent chromosomes from different origins. The dashed boxes on the vertical lines indicate 0 copy, while the solid boxes represent 1 copy. S06, S04, S07, S01, S05, S08, S02, and S03 represent the eight molecular marker loci, and the numbers in the graph indicate the PCR amplification fragment lengths (bp) of each molecular marker locus; (**b**) IGV plot of Family 4, showing the SMN1-FL haplotypes, with arrows indicating the inheritance of alleles across generations.

conclusion that the subject is a 2 + 0 carrier can only be drawn if the subject does not share any haplotypes with the parent. In cases of shared haplotypes, CASMA-trio can only provide probabilistic predictions without definitive answers. To improve the accuracy of predictions, accumulating frequencies of each specific haplotype from large populations may be beneficial. Additionally, in rare cases where parents have the same *SMN1*-FL haplotype, the results may be inconclusive.

In conclusion, CASMA has the capability to detect *SMN1* and *SMN2* copy numbers. Traditional STR linkage analysis has limitations in detecting *SMN1* 2 + 0 carriers due to the need for probands and participation of multiple generations. CASMA-trio, based on long-read sequencing and haplotype analysis, effectively addresses the shortcomings of STR linkage analysis and has significant application value in detecting *SMN1* 2 + 0 carriers. CASMA-trio has the potential to replace or augment the MLPA + STR linkage analysis approach, and its implementation will facilitate comprehensive screening of SMA carriers in a majority of cases.
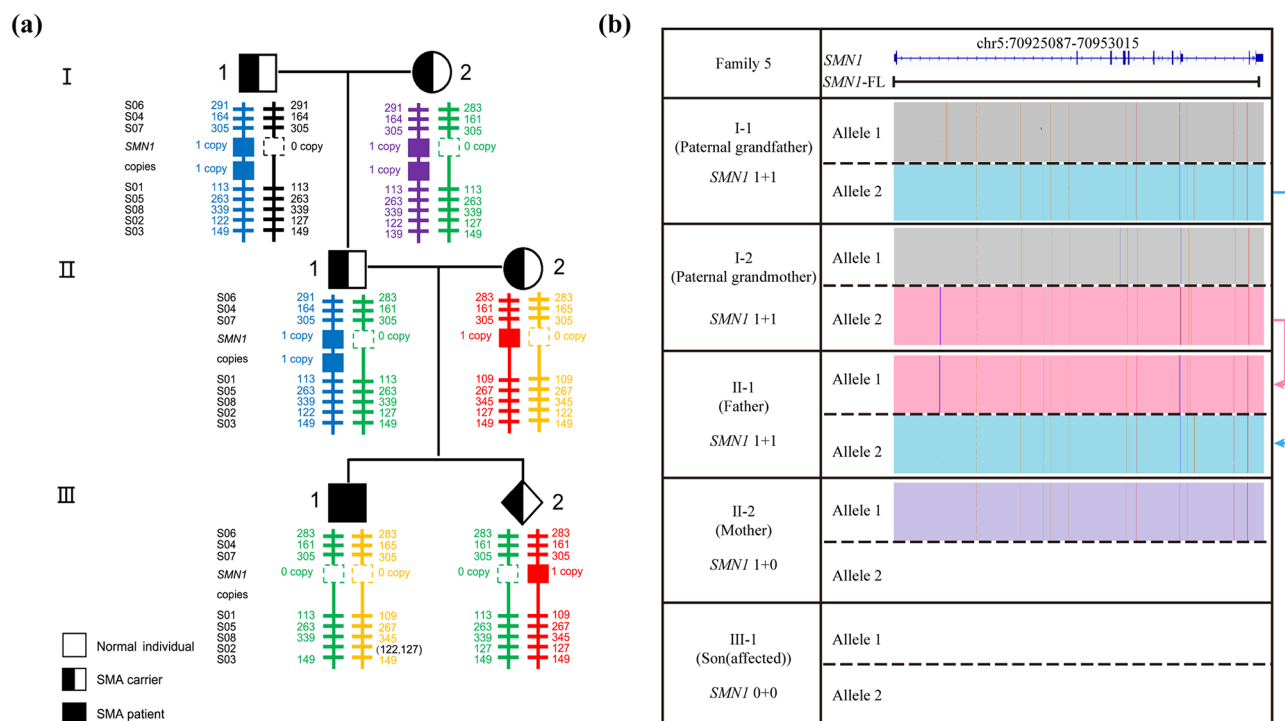
**Fig. 4.** Pattern graph for STR linkage analysis and CASMA-trio in family 5 (**a**) The genotypes of family members. The genotypes of mother and all the offspring were detected by STR linkage analysis. The genotypes of the father and paternal grandparents were also detected by STR linkage analysis but were speculated based on the genotypes of all the offspring. Vertical lines of different colors represent chromosomes from different origins. The dashed boxes on the vertical lines indicate 0 copy, while the solid boxes represent 1 copy. S06, S04, S07, S01, S05, S08, S02, and S03 represent the eight molecular marker loci. The numbers in the diagram indicate the PCR amplification fragment lengths (bp) for each molecular marker locus. Parentheses indicate that the haplotypes at this site cannot be distinguished; (**b**) The IGV plots shows the *SMN1*-FL haplotypes in family 5. Arrows indicate the inheritance of alleles across generations.

## Data availability
The datasets generated and analysed during the current study are available in the Open Science Framework repository, https://osf.io/fqhnk/?view_only=a82f856f1e014bb294022d25c3308844.

## References
1. Lunn, M. R. & Wang, C. H. Spinal muscular atrophy. *Lance* **371** (9630), 2120–2133 (2008).
2. Cao, Y. et al. Transmission characteristics of *SMN* from 227 spinal muscular atrophy core families in China. *J. Hum. Genet.* **65** (5), 469–473 (2020).
3. Chen, X. et al. Spinal muscular atrophy diagnosis and carrier screening from genome sequencing data. *Genet. Medicine: Official J. Am. Coll. Med. Genet.* **22** (5), 945–953 (2020).
4. Oskoui, M. et al. The changing natural history of spinal muscular atrophy type 1. *Neurology* **69** (20), 1931–1936 (2007).
5. Prior, T. W., Nagan, N., Sugarman, E. A., Batish, S. D. & Braastad, C. Technical standards and guidelines for spinal muscular atrophy testing. *Genet. Medicine: Official J. Am. Coll. Med. Genet.* **13** (7), 686–694 (2011).
6. Wurster, C. D. & Ludolph, A. C. Nusinersen for spinal muscular atrophy. *Ther. Adv. Neurol. Disord.* **11**, 1756285618754459 (2018).
7. Aartsma-Rus, A. FDA Approval of Nusinersen for spinal muscular atrophy makes 2016 the year of splice modulating oligonucleotides. *Nucleic acid Ther.* **27** (2), 67–69 (2017).
8. Gregg, A. R. et al. Screening for autosomal recessive and X-linked conditions during pregnancy and preconception: a practice resource of the American College of Medical Genetics and Genomics (ACMG). *Genet. Medicine: Official J. Am. Coll. Med. Genet.* **23** (10), 1793–1806 (2021).
9. Lefebvre, S. et al. Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* **80** (1), 155–165 (1995).
10. Sheng-Yuan, Z. et al. Molecular characterization of *SMN* copy number derived from carrier screening and from core families with SMA in a Chinese population. *Eur. J. Hum. Genetics: EJHG.* **18** (9), 978–984 (2010).
11. Zhao, M. et al. Identification of novel microsatellite markers flanking the *SMN1* and *SMN2* duplicated region and inclusion into a single-tube Tridecaplex Panel for Haplotype-based preimplantation genetic testing of spinal muscular atrophy. *Front. Genet.* **10**, 1105 (2019).
12. Wirth, B. An update of the mutation spectrum of the survival motor neuron gene (*SMN1*) in autosomal recessive spinal muscular atrophy (SMA). *Hum. Mutat.* **15** (3), 228–237 (2000).
13. Qu, Y. J. et al. Mutation spectrum of the Survival of Motor Neuron 1 and functional analysis of variants in Chinese spinal muscular atrophy. *J. Mol. Diagnostics: JMD.* **18** (5), 741–752 (2016).

14. Schouten, J. P. et al. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res.* **30**(12), e57 (2002).
15. Stuppia, L., Antonucci, I., Palka, G. & Gatta, V. Use of the MLPA assay in the molecular diagnosis of gene copy number alterations in human genetic diseases. *Int. J. Mol. Sci.* **13** (3), 3245–3276 (2012).
16. Lee, T. M. et al. Quantitative analysis of *SMN1* gene and estimation of *SMN1* deletion carrier frequency in Korean population based on real-time PCR. *J. Korean Med. Sci.* **19** (6), 870–873 (2004).
17. Feng, Y. et al. The next generation of population-based spinal muscular atrophy carrier screening: comprehensive pan-ethnic *SMN1* copy-number and sequence variant analysis by massively parallel sequencing. *Genet. Medicine: Official J. Am. Coll. Med. Genet.* **19** (8), 936–944 (2017).
18. Vidal-Folch, N. et al. Multiplex Droplet Digital PCR Method Applicable to Newborn Screening, Carrier Status, and Assessment of spinal muscular atrophy. *Clin. Chem.* **64** (12), 1753–1761 (2018).
19. Mailman, M. D. et al. Hybrids monosomal for human chromosome 5 reveal the presence of a spinal muscular atrophy (SMA) carrier with two *SMN1* copies on one chromosome. *Hum. Genet.* **108** (2), 109–115 (2001).
20. Alías, L. et al. Improving detection and genetic counseling in carriers of spinal muscular atrophy with two copies of the *SMN1* gene. *Clin. Genet.* **85** (5), 470–475 (2014).
21. Yanyan, C. et al. Familial study of spinal muscular atrophy carriers with *SMN1* (2 + 0) genotype. *Yi Chuan = Hereditas*. **43** (2), 160–168 (2021).
22. Li, S. et al. Comprehensive analysis of spinal muscular atrophy: *SMN1* Copy Number, Intragenic Mutation, and 2 + 0 carrier analysis by third-generation sequencing. *J. Mol. Diagnostics: JMD*. **24** (9), 1009–1020 (2022).
23. Li, S. et al. An effective and Universal Long-Read sequencing-based Approach for *SMN1* 2 + 0 Carrier Screening through Family Trio Analysis. *Clin. Chem.* **69** (11), 1295–1306 (2023).
24. Liang, Q. et al. A more Universal Approach to Comprehensive analysis of thalassemia alleles (CATSA). *J. Mol. Diagnostics: JMD*. **23** (9), 1195–1204 (2021).
25. Liu, Y. et al. Comprehensive Analysis of Congenital Adrenal hyperplasia using Long-Read sequencing. *Clin. Chem.* **68** (7), 927–939 (2022).
26. Liang, Q. et al. Comprehensive Analysis of Fragile X Syndrome: full characterization of the FMR1 locus by Long-Read Sequencing. *Clin. Chem.* **68** (12), 1529–1540 (2022).
27. Chen, K. L. et al. Duplications and de novo deletions of the *SMN*t gene demonstrated by fluorescence-based carrier testing for spinal muscular atrophy. *Am. J. Med. Genet.* **85** (5), 463–469 (1999).
28. Burlet, P. et al. Single-sperm analysis for recurrence risk assessment of spinal muscular atrophy. *Eur. J. Hum. Genetics: EJHG*. **18** (4), 505–508 (2010).
29. Rouzier, C., Chaussenot, A. & Paquis-Flucklinger, V. Molecular diagnosis and genetic counseling for spinal muscular atrophy (SMA). Archives de pediatrie: organe officiel de la Societe francaise de pediatrie, 27(7S), 7S9–7S14 (2020).

## Acknowledgements

## Author contributions

JCH and WZH conducted data analysis and wrote the manuscript. JJX, YCW, XYM, SRL, JFC, GYX, QQC collected blood samples and clinical data and performed DNA extraction. SYL and QL were responsible for the study design and guiding of the study implementation and revised the manuscript. All authors reviewed the manuscript.

## Funding

## Declarations

## Competing interests

The authors declare no competing interests.

## Ethics approval and consent to participate

This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of the Third Affiliated Hospital of Guangzhou Medical University (Ethics Approval [2021] No. 254). All patients provided their voluntary informed consent prior to the procedure being performed.

## Additional information

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-90603-1.

**Correspondence** and requests for materials should be addressed to Q.L. or S.L.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.