A Review on Diagnosis of Spinal Muscle Atrophy Using Copy Number Variation Method

Authors Details__ Carolin *EmadEldeen AbdElgaber, Martina Romany Mikhail, Kerolos Mamdouh Bakheet, Sherry Raafat Lamey, Abdelrhman Ahmed Mohamed, Mary Elias Naguib Elias, Mariam Nader Samuel Asham.*

I.ABSTRACT

This is a review of how the copy number variation (CNV) is used in the diagnosis of Spinal Muscular Atrophy (SMA). We will summarize six tools that measure the copy number of genes responsible for the normal or abnormal phenotype, these genes are the survival motor neuron 1 (SMN1) gene and SMN2.

The tools are: Code-Seq Technology, TaqMan Technology, SMAca Tool, PCR-RFLP, algorithm resolves the CNs of SMA1 and SMN2 Using short read GS data and novel integrated digital PCR.

Code-Seq Technology is a recently developed proprietary technology (patent pending) that utilizes the power of NGS to simultaneously analyze small nucleotide variations, copy number variations, and methylation status at multiple sites across the genome.

TaqMan Technology presents a new rapid and reliable approach to determine the copy numbers of the SMN1 and SMN2 genes. For the differentiation of the two genes, this tool presents a quantitative test based on TaqMan technology using minor groove binder (MGB) probes.

SMAca tool is the first python tool to detect SMA carriers and estimate the absolute SMN1 copy number using NGS data. Moreover, SMAca takes advantage of the knowledge of some variants specific to SMN1 duplication to also identify silent carriers.

The implemented PCR-RFLP (Polymerase Chain Reaction - Restriction Fragment Length Polymorphism), multiplex PCR, and real-time PCR is used to analyze the correlation between CNVs of SMN1, NAIP, and SMN2 genes and the SMA phenotype in samples taken from Egyptian patients. Three different molecular techniques are used to generate the genotype of SMA patients.

Fragment length polymorphism (PCR-RFLP) was used to detect the homozygous absence of exon 7 of the SMN1 gene using DraI restriction enzyme. Multiplex PCR amplification of exons 5 and

13 of the gene was used to analyze the deletion of the Neuronal Apoptosis Inhibitory Protein (NAIP) gene.

For the algorithm used to resolve the CNs of SMA1 and SMN2 Using short-read GS data, the recent advances in next-generation sequencing (NGS) made it possible to develop a method that accurately identifies the CN of SMN1 and SMN2 using genome sequencing data by analyzing read depth and eight informative reference genome differences between SMN1/2.

Lastly, for the novel integrated digital PCR instrument, the digital PCR (dPCR) technology has been proven to be highly sensitive and accurate in detecting copy number variations (CNV). Here, a rapid multiplex SMA dPCR genotyping assay was run on a fully integrated dPCR instrument with five optical channels. The hydrolysis probe-based multiplex dPCR assay quantifies the genes SMN1, SMN2, and the total SMN (SMN1 + SMN2) while using the RPPH1 gene as an internal reference control.

Keywords— Copy number variation, Copy number analysis, Spinal muscle atrophy, SMN1, Bioinformatics.

II. INTRODUCTION

Spinal muscular atrophy (SMA) is a group of genetic disorders that injure the motor neurons inside the brain stem cells and the spinal cord that regulate the activity of primary skeletal muscle, which yields muscle weakness and atrophy. SMA is classified into four types. Type 1 SMA, which is known to be the most severe type, causes severe muscle weakness presenting before 6 months of age with a median survival of 24 months [2].

Type 2 SMA starts between 6 and 12 months of age. Patients with type 2 SMA can sit vertically. However, they are incapable of walking independently.

Type 3 SMA, also known as juvenile SMA, is often distinguished by the onset of symptoms after 18 months after birth. Patients with type 3 SMA can walk and have a steadier degree of disease severity, and more than half of them remain mobile after 40 years of age [1, 3].

Type 4 SMA is an adult-onset type with the mildest clinical severity and is associated with normal life yearning [1]. Type '0' SMA is characterized by prenatal onset resulting in severe birth hypotonia [1].

The gene SMN1 is located at the chromosomal 5q13 locus and the highly homologous SMN2 gene (SMN2, OMIM*601627) is located close to SMN1.

Most of SMA cases (95-98%) are due to homozygous deletion of exon 7 of the SMN1 gene (SMN1, OMIM*600354). The remaining 2-5% cases result from the presence of small sequence variants in the SMN1 gene present in the compound heterozygous state with SMN1 exon 7 deletion [4].

Most of SMA carriers have one functional copy of the SMN1 gene on only one of their chromosomes. While in a relatively small percentage of individuals (3.3-8.5% depending on ethnicity) there are two functional copies of the SMN1 gene on one chromosome in the cis-state and zero or no copies on the other chromosome [5]. Consequently, although they have normal

functional SMN1 '2' genes, they still behave like carriers. This phenomenon is known as the “2+0” genotype and individuals who has this genotype are known as “silent SMA carriers” [5, 6]. Copy number difference is related to the degree of SMA.

Copy number variation (CNV) refers to a circumstance in which the copy number of a specific DNA segment varies between the genomes of different individuals.

These variants may be short or may include thousands of bases. These structural differences usually occur due to duplications, deletions, or other types of mutations that can affect long segments of DNA. These regions may or may not contain a gene(s).

This survey reviews six tools used in copy number analysis for SMA which are CODE-SEQ technology, SMAca Tool, TaqMan Technology, PCR-RFLP (multiplex PCR and real-time PCR), algorithm resolve the CNs of SMN1 and SMN2 using short-read GS data and novel integrated digital PCR instrument.

III.PROBLEM DEFINITION

Spinal muscular atrophy (SMA) is a genetic disease that affects the peripheral nervous system, central nervous system, and voluntary muscle movement, which is skeletal muscle, that causes muscle weakness and atrophy as muscles get smaller.
SMA affects a child's ability to crawl, sit up, walk, and control head movements in general. Severe SMA may lead to death as it causes damage to the muscles used for breathing and swallowing. SMA is also the leading genetic cause of infant death after cystic fibrosis [16].
There are four types of SMA (type I, type II ,type III and type IV). These types range from 6 months old to 35 years old.
The incidence of SMA is 1 in 6000–10,000 live births, and the carrier frequency is 1:40–80 among different ethnic groups [17].
One of the biggest problems facing SMA patients is the late diagnosis of the disease. Late diagnosis lowers the chance of treatment and improvement.
While it is proved that early detection of SMA can essential for long term quality of life according to the availability of two early treatments, nusinersen [20] and Zolgensma [21], which have received FDA approval for the improvement of SMA symptoms.
SMA is caused by a deficiency of a motor neuron protein called survival of motor neuron (SMN). This protein is vital for normal motor neuron function.
SMN plays a pivotal role in gene expression in motor neurons. Its deficiency is caused by mutations on chromosome 5 in a gene called SMN1 which codes for SMN protein. The most common alteration in the SMN1 gene within patients diagnosed with SMA is a deletion of a whole segment, called exon 7 [14].
This deletion case is present in 94% of all SMA cases. Neighboring SMN2 genes can in part compensate for nonfunctional SMN1 genes as there is 99% identity between these two genes [15]. The sequence similarity between the two genes SMN1 and SMN2, and the of the SMN locus complexity makes the identification of the copy number of SMN1 by next-generation sequencing (NGS) very difficult. Another obstacle is that the tests cannot detect that some people are carriers. This is known as silent carriers.

IV.BACKGROUND

This survey aims to review methods of diagnosis of SMA that are centered around the analysis of copy number variation (CNV) in the genome sequence data.

This survey reviews six scientific papers and the CNV method is used in each. We use the copy number method to determine whether the patient is a carrier of the disease or not using genome sequencing (GS) data by the analysis of read depth and eight informative reference genome differences between SMN1/2 [22].

demonstrated to us a mysterious concept, which is the silent carrier, sometimes people have more than 2 altered copies of cis-SMN1 on the same chromosome.

People with this structure are known as silent carriers because the tests cannot detect that they are a carrier. This paper demonstrates how to diagnose muscular dystrophy patients and specify the silent carrier [23].

Various molecular techniques such as restriction fragment length polymorphisms (RFLP), quantitative polymerase chain reactions (qPCR), next-generation sequencing (NGS), and digital droplet PCR (ddPCR) have been used in the diagnosis for SMA carriers [24][10].

Gold standard for the detection of heterozygous SMN1 is the multiplex ligation-dependent probe amplification (MLPA) technique (MRC-Holland, Netherlands) [25, 26]. For all its advantages, MLPA has some constraints, especially when it comes to population-level screening, because it has a narrow ability to be multiplexed, demands several manual processing steps, and interpretation of results can be challenging.

A recently developed proprietary technology platform “CODE-SEQ” (patent pending) uses the power of NGS to analyze small nucleotide variations, copy number variations, and methylation status at numerous sites across the genome [24].

We also knew that there is another modifier gene with SMN2 whose copy number can affect the severity of SMA. We concluded that the combination of modifier genes to provide the prognostic genetic pattern for phenotype determination is preferable to using Copy number variations (CNVs) of exon 7 of SMN2 gene only. CNVs of exon 7 of SMN2 are of high importance to predict patients’ responses to genetic therapy.

On the other hand, deletion of exon5 of NAIP gene alone is not a sufficient predictor of SMA severity as the Copy number variations (CNVs) of exon 7 of SMN2 gene significantly decrease with the increase in disease severity [27].

MGB probes form extremely stable duplicates with single-stranded DNA targets, allowing shorter probes to be used for hybridization-based assays. MGB probes have higher melting temperatures and are consequently more sensitive to single base mismatches than conventional probes, we will explain later how important MGB probes are to distinguish SMN1 from SMN2[13].

V.RELATED WORK

This section summarizes how each tool was applied and what methodology and sequence of steps were used to analyze CNV in the SMN1 gene and SMN2 gene.

5.1CODE-SEQ technology [24].

Validation studies for the CODE-SEQ on pre-characterized reference samples pre-characterized reference samples with known copy numbers in SMN1 and SMN2 and single nucleotide variations (SNVs) in the SMN1 gene associated with the "2+0" genotype were obtained from NIGMS Human Genetic Cell Repository at the Coriell Institute for Medical Research.

The SALSA® MLPA® probemix P460 was used to confirm the copy numbers of SMN1 and SMN2, as well as the presence or absence of both SNVs associated with the "2+0" genotype.

Each of the reference samples, along with eighteen pre-characterized human samples, was analysed using the CODE-SEQ platform in duplicate.

The results of CODE-SEQ analysis were compared with the results from the gold standard MLPA assay.

The cutoffs for calling out copy numbers (from "0" to "4") were estimated.

Blinded clinical samples A total of 80 clinical samples (blinded to the operators and analysis team) along with six reference samples were analysed using the CODE-SEQ.

All clinical samples were previously evaluated to determine CNs of exons 7 and 8 of the SMN1 and SMN2 genes by MLPA probemix P060.

Multiplex ligation-dependent probe amplification analysis Multiplex ligation-dependent probe amplification (MLPA) was done as recommended by the manufacturer.

The SALSA® MLPA® probemix P060 was used for the detection of copy numbers of exons 7 and 8 of SMN1 and SMN2.

MLPA probemix P460 was used to further confirm the copy numbers in SMN1 (exons 7 and 8) and SMN2 (exon 7) genes and the presence or absence of "2+0" genotype-associated SNVs.

This tool did not include any clinical samples known to harbor SNVs associated with the "2+0" genotype, but reference samples with this configuration were tested.

In brief, the protocol involves the hybridization of the sample DNA with the premade pool of eighteen pairs of "CODE-SEQ SMA ProbeMix v8.0" coded oligonucleotides.

Each of the oligonucleotide pairs is coupled with a unique probe identification sequence (PIDS) and targets a specific region in the human genome.

Of these 18 pairs of coded oligonucleotides, five target SMA-related loci, and 13 are reference oligonucleotide pairs targeting the unlinked human genomic regions (which were used for data normalization).

Of the five SMA-targeted oligonucleotide pairs, two pairs are specific to single nucleotide variations (SNVs) that differentiate between the SMN1 gene from the SMN2 gene.

For the detection of haplotypes associated with the "2+0" genotype, the two SNVs targeted were g.27706-27707delAT and g.27134T>G.

Post-hybridization, sample-specific identifiers (SIDS) were integrated using PCR to enable a multiplexed analysis of samples by next-generation sequencing (NGS).

Analysis was done using GenePath Dx's CODE-SEQ Dx (v1.02) analytics platform.

In brief, the steps involve intra-sample normalization, which enables a tolerance for a wide dynamic range of the sample input, and inter-sample normalization, which provides a robust baseline.

CODESEQ analytics platform v1.02 then ascribes results for each sample based on a rule-based methodology with logical operators such as OR, AND, NOT, and XOR, generating a "normalized read ratio" (NRR) value for each oligonucleotide pair targeting SMN1, SMN2, or unlinked human genomic regions.

Calculation of estimated CNs for genomic targets Copy numbers for each genomic target were inferred based on the computed NRR values.

Also, manual analysis of raw sequence reads was performed in cases where (i) there was a discrepancy between estimated copy numbers (ECNs) of exon 7 and exon 8 of the SMN1 gene, (ii) the NRR value was at the borderline or between the pre-defined cutoffs for copy number calculations, or (iii) the results were suggestive of the presence of any of the SNVs associated with the "2+0" genotype.

The clinical interpretation of the data was performed based on the ECNs of oligonucleotide pairs targeting the SMN1 gene and the presence or absence of SNVs associated with the "2+0" genotype.
Homozygous deletions of SMN1 exon 7 were considered consistent with a diagnosis of SMA.
Heterozygous deletions of SMN1 exon 7 were considered consistent with a diagnosis of the individual being a heterozygous carrier for SMA.
A result of more than "2" copies of the SMN1 gene on exon 7 was designated as normal, with a very low residual risk of the individual being an SMA carrier; where the read counts of oligonucleotide pairs targeting the two SNVs associated with the "2+0" genotype were also checked.
If one or both of these SNVs are detected in a sample, then there is an increased chance of the individual being a silent carrier.
In cases where none of the SNVs were present in a sample, the individual was considered to have a low risk of being an SMA carrier, and partner screening wouldn't be recommended.

5.2 SMAca Tool [28]

1.NGS data processing

Data were generated by a NovaSeq 6000.
Raw FASTQ files were processed by following a standard NGS pipeline, and data were generated using paired-end reads (2 × 150 bp). In brief, after filtering out low-quality reads with fast v0.20.0 [28], reads were aligned with BWA–MEM v0.7.16 [29] against the human reference genome GRCh37/h19. Any potential polymerase chain reaction duplicates were marked with Picard v2.17.3.
At last, the whole set of BAM files was analyzed with SMAca in a single batch. A set of 1109 alignment CRAM files, comprising a diverse set of individuals from multiple populations, were downloaded from the 1000 genomes data portal. Moreover, a set of 1109 alignment CRAM files, comprising a diverse set of individuals from multiple populations, were downloaded from the 1000 genomes data portal.
The corresponding SMN1 and SMN2 validated copy number (CN) statuses were obtained from Vijzelaar et al. The whole set of CRAM files was analyzed with SMAca in a single batch. Samples with experimentally validated one SMN1 gene copy predicted to be SMA carriers were marked as true positives.

2.SMN1 copy-number estimation

SMAca fisrt calculates the raw proportion of SMN1 reads over the total number of reads covering SMN1 and SMN2 at three specific gene positions (marked as a, b, and c) for each sample (Table1)
These positions resemble single nucleotide differences between SMN1 and SMN2. Raw values are then scaled concerning 20 control genes (Table 2) previously described to have consistent average coverage relative to SMN1 and SNM2. Besides, two genetic variants that have been associated with duplication events in SMN1 are also screened and reported.

TABLE 2 Control genes

Note: List of genes used to calculate the scale factor (θ ^ i) and the scaled proportion of SMN1 reads (πij).

ACAD9	*FASTKD2*	*ITGA6*	*NTRK1*	*SIL1*
ATR	*FOXN1*	*IVD*	*PTEN*	*SLC22A5*
CYP11B1	*HEXB*	*LMNA*	*RAB3GAP1*	*SLC35D1*
EDNRB	*IQCB1*	*LRPPRC*	*RAPSN*	*STIM1*

The relative coverage of SMN1 and SMN2 with respect to each control gene is calculated: $Z_{ki} = (c_{i1} + c_{i2})/H_{ki}$, were c_{i1} and c_{i2} are the average coverage for the SMN1 and SMN2, and H_{ki} represents the average coverage for the control gene k in the ith sample. Thus, the scale factor $\hat{\theta}_i = (\sum_{k=1}^{K} Z_{ki} / \overline{Z}_{ki})/K$, where $\overline{Z}_{ki} = \sum_{I=1}^{N} Z_{ki})/N$, N is the total number of samples and K the total number of control genes, is calculated for each sample. At last, the raw proportion of SMN1 reads are scaled: $\pi_{ij} = \hat{\theta} \times D_{1ij}/(D_{1ij} + D_{2ij})$, where D_{1ij} and D_{2ij} are the raw coverage for SMN1 and SMN2 at position j in the ith sample.

1. SMA carrier categorization
The classification of SMA carriers according to some rules that are used to estimate the absolute copy-number as follows:
• Genotypes 1 SMN1:3 SMN2 are expected to have $\hat{\theta}$ i ~ 1 and D_1ij/D_2ij~ ⅓.
• Genotypes 1 SMN1:2 SMN2 are expected to have $\hat{\theta}$ i ~ 0.75 and D_1ij/D_2ij ~ ½.
• And genotypes 1 SMN1:1 SMN2 are expected to have $\hat{\theta}$ i ~ 0.5 and D_1ij/D_2ij~ 1.
For the silent carriers detection, samples with two polymorphisms (g.27134T>G and g.27706_27707delAT) associated with duplication events in SMN1[31] are selected. According to the SMN2 copy number, the expected $\hat{\theta}_i$ is supposed to be close to 0.75 (2:1) or 0.5 (2:0).
The scaled coverage proportion of the SMN1 gene in both cases is often close to ½ in each position.
SMA carriers can be detected either in genome or exome sequences, and even small panels as long as SMN locus and the control genes are covered. However, for silent carriers, which require the analysis of some intronic positions, the use of genomic sequences is highly recommended.

5.2 TaqMan Technology [29]

Genomic DNA was collected from peripheral blood of 40 SMA patients carrying a homozygous SMN1 deletion (in the following called ‘‘SMA patients’’), 40 unrelated parents of SMA patients, and 100 normal controls from the general population. DNA was isolated from blood samples by a simple salting-out procedure [Miller et al., 1988] or in external laboratories, photometric measuring of DNA was performed on a Gene Quant II (Pharmacia, Erlangen, Germany). DNAs were diluted in water to a concentration of 20 ng/ml and stored at room temperature overnight before use.

The probands were typed for the multicopy markers AG1-CA and C212 [for details see Scheffer et al., 2001] in the SMA region. To estimate the influence of cross hybridizations between the SMN1 and SMN2 genes and the respective probes, DNA samples of four patients each with homozygous SMN1 or SMN2 deletions, respectively, were analyzed for the respective probes. These patients carried between one and four alleles of the multicopy markers AG1-CA and C212. The same DNA

sample of a normal male individual was used as standard in all amplification reactions for SMN1 and SMN2.

The presence of two SMN1 and SMN2 copies in this control was determined by analysis of the aforementioned multicopy markers revealing four alleles in the proband. In case of SMA patients, we used a SMA patient with two SMN2 copies as an additional standard. We used MGB probes to distinguish between SMN1 and SMN2; probes and primers were designed using the PrimerExpresst software (Applied Biosystems, Darmstadt, Germany).

The MGB probes were specific for the SMN1 and the SMN2 copies at position 6 in exon 7. The two probes contained a FAM reporter dye connected to the 50 end. Probes and primers were purchased from Applied Biosystems. As reference locus we used exon 3 of factor VIII: primer and probe information was published by Wilke et al. [2000].

PCR was carried out using an ABI Prism 7000 sequence detection system and 96-well MicroAmp optical plates (Applied Biosystems). The PCR was performed in a total volume of 25 ml, containing 50 ng of genomic DNA, 300 nM of each primer, 13 ml PlatinumsqPCR Supermix-UDG (Invitrogen, Karlsruhe, Germany), 0.5 mM ROX as passive reference (Invitrogen), 2 mM MgCl2, and 100 nmol of each MGB probe.

All of the reactions of the same run were prepared from the same master mix. Each well of the 96-well plate contained either 50 ng sample DNA, or 125 ng, 25 ng, or 5 ng standard DNA, respectively. Each of the test sample and each amount of standard DNA was amplified in two different wells. Reactions for the SMN1or SMN2 test loci and the factor VIII reference locus were prepared and run in parallel. PCR conditions were 2 min 501C, 10 min 951C, 40 cycles consisting of 15 sec 951C, and 1 min 601C.

The data evaluation was carried out using the ABI7000SDS software as described by Wilke et al. [2000]: separate standard curves were generated for the test loci and the reference loci. Using these curves, the starting gene copy number relative to the reference individual was determined for each well, and the mean of the relative starting gene CN values in different wells was calculated for each test sample and each locus. For male individuals, the absolute copy number of the SMN1 test locus per haploid genomic equivalent was calculated by the ratio of the means of the relative starting copy numbers of the test locus and the factor VIII reference locus. For female individuals, this ratio was multiplied by a factor of two.

For the diploid copy number, the values were multiplied with a factor of two.

5.3 PCR-RFLP (multiplex PCR and real-time PCR) [27]

Forty-four patients were referred from Clinical Genetics outpatient clinics. SMA diagnosis was according to the criteria specified by the International SMA Consortium.

Ten healthy individuals were used as a control group, for the normalization of the real-time quantification data results within each run. Written informed consent was taken from all subjects or their guardians.
Genomic DNA was extracted from the peripheral blood of SMA patients and controls according to the standard methods. All patients had a homozygous deletion of exon 7 of the SMN1 gene, which was determined by PCR-RFLP. Both exon 5 deletions of the NAIP gene and copy number variations of exon 7 of the SMN2 gene were also estimated.

Deletions within SMN1 and NAIP genes

The homozygous absence of exon 7 of the SMN1 gene was detected using restriction fragment length polymorphism (PCR-RFLP) using the DraI restriction enzyme. Deletion of the Neuronal Apoptosis Inhibitory Protein (NAIP) gene was analyzed using multiplex PCR amplification of exons 5 and 13 of the gene.

SMN2 genes dosage analysis

Copy number variations (CNVs) of exon 7 of the SMN2 gene were determined by quantitative real-time PCR on Step One Real Time PCR System (Applied Biosystems, USA) using primers designed to distinguish between SMN1 and SMN2 according to Feldkötter and Schwarzer. The PCR reaction was done in a final volume of 15 ul in duplicates containing 7 ul SYBR Green I master mix (Qiagen, Germany), 10 ng of DNA and 1 pmol of each primer. A melting curve was performed after each reaction to confirm proper amplification.
The calibration curve was constructed using Human Genomic DNA (Promega, Germany) by 2-fold dilutions (1.25, 2.5, 5, and 10 ng) representing 1, 2, 4, and 8 copies of the SMN2 gene, respectively.
Determination of CNVs of exon 7 of the SMN2 gene was performed by plotting the Ct (cycle threshold) value of each sample on the calibration curve to calculate CNVs of exon 7 of the SMN2 gene.

Statistical analysis

Kruskal-Wallis test was performed to infer the relationship between CNV of exon 7 of SMN2, the deletion in exon 5 of the NAIP gene, and patients' phenotype. Statistical analysis was done using SPSS software.

5.4 Algorithm resolve the CNs of SMN1 and SMN2 using short-read GS data [22].

5.4.1 Samples and data processing:

Samples analysed and confirmed using digital PCR were procured from the Motor Neuron Diseases Research Laboratory.
Historical patient samples with known SMA or carrier condition measured by MLPA, GS was performed on samples with digital PCR results, samples with MLPA results, and the population samples from the Genomes Project.
The sequencing and processing of this data were done using different kinds of sample preparation methods, Illumina sequencers, and read aligners.

5.4.2 SMN copy-number analysis by orthogonal methods:

SMN1 and SMN2 copy numbers were measured for samples using the QuantStudio 3D Digital PCR System (Life Technologies) using allele-specific exon 7 probes as described previously. SMN1 and SMN2 copy numbers were normalized against those for RPPH1, historical patient samples were previously tested in a clinically accredited laboratory by the MLPA for SMN1/2 exons 7 and 8 copy number. Moreover, two samples from the Next Generation Children project were confirmed using MLPA.

5.4.3 Copy-number calling for intact and truncated SMN:

Two common copy-number variants (CNVs) involve the loci of SMN1 and SMN2 , the whole-gene CNV, and a partial gene deletion of exons 7 and 8 (SMN2Δ7–8), the tool first counts reads that align to either SMN1 or SMN2.
Read counts in a 22.2-kb region containing exon 1 to exon 6 are used to calculate the total SMN and CN, and read counts in the 6.3-kb region including exons 7 and 8 are used to estimate the CN of intact
SMN.

5.4.4 Genotyping SMN1/2 copy number using differentiating bases:

This tool calls the number of chromosomes carrying the SMN1 and SMN2 genes' bases by combining the total SMN CN with the read counts supporting each of the gene-specific bases. At each SMN1/2 differentiating base, based on the called CN of intact SMN, the caller iterates through all possible combinations of SMN1 and SMN2 CNs and derives the combination that produces the highest posterior probability for the observed number of SMN1 and SMN2 supporting reads.
The SMN1 CNs called at single bases are then merged to make the aggregate SMN1 CN call based on a consensus rule.
This method can be applied to different positions to identify the copy number of bases known to be specific to one of the two genes.

5.5 Novel integrated digital PCR instrument [10]:

The Absolute Q digital PCR System has a walk-away workflow similar to traditional qPCR. The MAP plate is loaded through pipette with 10 μL of PCR mix and an overlay of 10 μL isolation Buffer in each well. The wells are then covered with specialized gaskets. The plate is then placed into the Absolute Q tray and retracted inside the system.
The Absolute Q employs positive pressure from an on-board compressor to divide the sample within the MAP plate without the microfluidic valves, sealing films or other moving parts. The MAP slides are made of a cyclo olefin polymer (COP; 80 μm) film that seals the microfluidic components that are molded into a separate piece of thicker material.
Four similar slides are bonded to a rigid, microtiter format plate frame that includes the loading wells to complete the plate assembly.
The thin film becomes gas permeable when positive pressure is applied to a well including the reagents. As the reagent enters the microfluidic features, air is passed out of the partition via film. This allows reagent to fully fill dead-ended partitions and prevents any bubbles from forming inside of the microfluidic features.
The Isolation Buffer overlay follows the reagent and physically separates the reagent reaction volumes to satisfy the partitioning. Positive pressure used to the consumable during the PCR thermocycling prevents any evaporation and guarantees that bubbles will not form and disrupt the isolated micro-reactions.
Before and after PCR thermocycling, the entire arrays are imaged with up to five optical channels configured for the most used dyes, including a ubiquitous quality control dye (ROX) used to confirm proper partition filling and finding.
The images taken before PCR are subtracted from the after-PCR images to remove any background noise. Combination Analysis software automatically applies for optical crosstalk compensation and categorizes the partitions using a convoluted neural network algorithm to eliminate false positives/ negatives and validates the results so that it would be a robust quantification.
The full PCR process occurs within the single benchtop instrument without the operator's interaction after setting up the protocol parameters. The streamlined workflow decreases the

potential for contamination, minimizes human handling errors, and reduces the time to result. Fluids never contacted the instrument components, so minimal system maintenance is required.
For Digital PCR, all the control DNA samples were diluted to 10 ng/µL working stock. The blinded DNA samples were calculated using a Qubit 4 fluorometer (Invitrogen) and diluted to 10 ng/µL.
The PCR reaction mix contained 1X Combination MasterMix, 2.5–25 ng human gDNA, 900 nmol/L RPPH1 primers, 1800 nmol/L SMN1/SMN2 primers, 900 nmol/L total SMN primers, and 250 nmol/L of each probe.
Each MAP plate well was loaded by hand through pipette with 10 µL of PCR reaction mix then overlaid with 10 µL Isolation Buffer. The gasket caps were placed, and the plate was put into the Absolute Q and run with the following conditions: 3 minutes activation at 95 °C, 40 cycles of 5 s at 95 °C, and 30 s at 62 °C. The automated Absolute Q Control Software (v1.0.20) controls the sample digitization into the MAP partitions, imaging and thermal cycling.
For Data analysis, the copy numbers were estimated using the Absolute Q Analysis Software (v10.5.4). The software automatically carries out the optimal positive/ negative threshold and in this case, the positive signals were five times more elevated than the negative signals. For each unit, the software shows the total viable partition count and the positive partition count.
The software has a setting to identify sample types as a copy number variation assay and applies the suitable calculations to automatically display the copy numbers for the samples. The copy numbers of SMN1, SMN2, and total SMN are normalized by the reference control RPPH1 by the following equations:
SMN2CN=2 [ln(N/N−N2)/ln(N/N−N4)]
SMN1CN= 2 [ln(N/N−N1)/ln(N/N−N4)]
Total SMNCN=2 [ln(N/N−N3)/ln(N/N−N4)]
N = total number of viable partitions, N1 = number of SMN1 positive partitions, N2 = number of SMN2 positive partitions, N3 = number of total SMN positive partitions and N4 = number of RPPH1 positive partitions.
For Statistical analysis, the coefficient of variability (%CV), defined as the standard deviation divided by the mean value for each set of replicates, was used to evaluate repeatability or intra-assay precision. Reliability was also measured by the intraclass correlation coefficient (ICC) using the SPSS v.25 (IBM, Armonk, NY). A Bland Altman agreement analysis45 using the SigmaPlot v.12.0 (Systat Software, Inc., San Jose, CA) was used to measure the agreement between the dPCR methods.

VI. EXPERIMENTAL RESULTS

6.1 CODE-SEQ technology [24]:

Validation results of CODE-SEQ with reference samples.
Eight reference samples were analyzed by the MLPA (ProbeMixes P060 and P460) and SMA CODE-SEQ (ProbeMix v8.0).
Analysis revealed that in 39 out of the 40 (97.5%) data points being evaluated, the results of CODE-SEQ matched the expected results.
One target (exon 8 of the SMN1 gene) in one sample (HG07) was a mismatch.
For this particular target, the copy number determined by the CODE-SEQ assay was 3, while it was expected to be 2.

On succeeding manual analysis, it was noted that the NRR value for this data point from the CODE-SEQ experiment was at the borderline (1.34) for the binning values for two copies versus three copies.
Moreover, when the CODE-SEQ results were compared with the results from the MLPA P460 assay, the copy numbers for 37 of the 40 data points (92.5%) matched.
Concerning the two SNVs (g. 27134T>G and g. 27706-27707delAT), CODE-SEQ correctly witnessed the presence of both markers in all 5 of the reference samples that harbored them.
Interestingly, the results of the MLPA P060 and P460 assays didn't correlate for 3 of the 24 common data points.
Results of 8 reference samples processed in duplicate by SMA CODE-SEQ Each of these eight reference samples was run in duplicate.
The blinded study included 80 clinically characterized samples and 6 reference DNA samples from the Coriell Institute.
94 % of the total reads that passed quality filters were mapped to targeted genomic regions, while only 0.2% showed cross-mapping to non-targeted regions of human genome.
18 of 80 clinical samples processed (22.5%) showed the existence of homozygous deletions in exon 7 of the SMN1 gene and were therefore labeled as "affected with SMA."
Twenty-one samples of 80 (26.2%) were detected with a heterozygous deletion in exon 7 of the SMN1 gene and were classified as SMA carriers.
Although of limited clinical significance, to benchmark assay performance, the copy number results were compared for the SMN2 gene in addition to those of SMN1.
The quantitative detection matched the known copy number in six Coriell reference samples.
Among the 80 clinical samples tested in this research, 39 (48.75%) were diploid, 22 (27.5%) harbored heterozygous deletion, 16 (20%) harbored heterozygous duplication, two (2.5%) harbored homozygous deletion, and one sample only harbored homozygous duplication for the SMN2 gene.
The CODE-SEQ results correlated with the MLPA results for all 80 samples, delivering 100% correlation.
In addition to the SMN1 and SMN2 genes copy number detection, the platform was correctly able to identify the presence of SNVs associated with the "2+0" genotype in the reference samples that were tested.

6.2 SMAca Tool [28]:

The 1000 Genomes Project NAGEN1000 was used to screen a dataset of 326 genomes. Among them, seven samples (2.15%) were identified and successfully validated as putative SMA carriers and successfully validated by the MLPA (Table S1a–d). The percentage of the predicted and confirmed SMA carriers in our dataset fits perfectly to the expected carrier frequency (2.10%) previously described in the bibliography. The genotype calculations for the SMA carrier samples agreed with the experimental validation as shown in Table 3 (except for case number 7 where the genotype could not be estimated). Also, case number 1 corresponds to an SMA carrier with an extra copy of SMN exons 1–6 (SMN1/2Δ7-8).

TABLE 1 SMN1 and SMN2 different nucleotides

Position	SMN1	SMN2
a	chr5:70247724	chr5:69372304
b	chr5:70247773	chr5:69372353
c	chr5:70247921	chr5:69372501

Note: Positions in SMN1 gene (and the analogous positions in SMN2) used to estimate the raw proportion of SMN1 reads (D1ij) over the total read number covering SMN1 and SMN2 (D1ij + D2ij).

This tool also performed a large-scale validation of 1109 genomes from a various set of individuals from multiple populations. Results show an overall high accuracy (0.998) and an F1 score of 0.938 (see Table S1 with the validation test, Table S2 with the full SMAca output, and Table S3 with the list of validated samples).

TABLE 3 MLPA results

No. of id	PI_a	PI_b	PI_c	cov SMN1a	Cov SMN1b_e7	Cov SMN1c	Cov SMN2a	Cov SMN2b e7	Cov SMN2c	Scale factor	CN estimation	MLPA genotype
1	0.26	0.19	0.24	13	8	11	23	23	22	0.741	1:2	1:1*
2	0.29	0.22	0.29	17	14	21	46	54	57	1.101	1:3	1:3
3	0.22	0.22	0.21	15	15	13	42	41	38	0.842	1:2	1:2
4	0.32	0.26	0.23	27	18	13	42	39	33	0.837	1:2	1:2
5	0.28	0.26	0.22	25	24	20	72	78	79	1.109	1:3	1:3
6	0.30	0.26	0.25	23	22	19	65	74	68	1.148	1:3	1:3
7	0.40	0.32	0.30	23	19	18	30	36	37	0.936	Inconclusive	1:2

Note: PI_x: scaled proportion of the SMN1 gene reads in position x; cov xp: raw coverage of gene x at position p; scale factor: θ ^ i; CN estimation: absolute copy number estimation SMN1:SMN2; MLPA genotype: genotype inferred from MLPA analysis. The MLPA analysis showed deletion of exons 7–8 on both genes but three copies of exons 1–6 (impossible to distinguish whether they come from SMN1 or SMN2).

With the idea of facilitating the introduction of the SMAca in production NGS pipelines, it has been optimized for running in different computer environments. the special stress has been made in the parallelization for exploiting multiple cores/processors when available. Figure 1 below shows the runtimes with an increasing number of processors. The measure of SMA mutational and copy number status for 326 genomes from Navarra 1000 Genomes Project NAGEN1000 takes almost 1 h and a half in one core but can be reduced to only 3 min in 24 cores (see Figure 1).

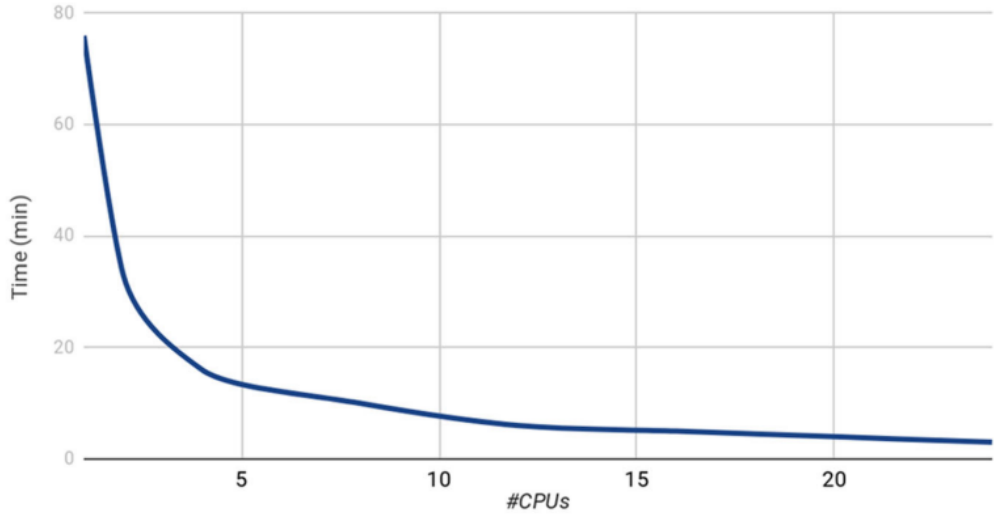

FIGURE 1 SMAca performance. Elapsed time for the analysis of the dataset (326 whole genome sequences) and different numbers of CPUs. The analysis of the whole dataset takes only 3 min by using 24 threads.

6.3 TaqMan Technology [29]:

The reproducibility of the quantification approaches for exons 7 of the SMN1 and SMN2 genes were determined in five independent PCR reactions of the same DNA samples on different days. In case of SMN1, we used four DNAs from male and female carriers and control persons, respectively. The range of measured SMN1 dosage was between 0.74 and 1.06 for the carriers, and between 1.68 and 2.16 for the control persons. To estimate the reproducibility of the SMN2 quantification approach, we tested four DNAs from SMA patients carrying between one and four SMN2 copies; in these individuals, the putative number of SMN2 copies was estimated by haplotype analysis of the Ag1-CA and C212 marker alleles. The measured SMN2 copy numbers ranged from 0.98–1.10 in the carrier of one copy, from 2.00 to 2.16 in the carrier of two copies, from 2.64 to 3.04 in the patient with three copies, and from 3.63 to 4.34 in case with four copies. To rule out that cross-hybridizations of the SMN1 specific MGB probe with the SMN2 copy influenced the real-time PCR efficiency, four patients with homozygous SMN1 deletions and different numbers of AG1 and C212 alleles corresponding to different SMN2 copy numbers were investigated. These different DNA samples failed to detect amplification, thereby proving specific hybridization of the probe to the SMN1 gene. For SMN2, a faint cross-hybridization of the SMN2-specific MGB probe with the SMN1 gene could be observed by analyzing patients homozygous for the SMN2 deletion.

While the results of the SMN2 copy numbers of the controls were not influenced, the mean measured SMN2 copy number was decreased in SMA patients when the standard DNA with two SMN1 copies was used for quantification. Taking into account this weak cross-hybridization, we used a DNA sample of an SMA patient as the standard to analyze the SMN2 copy number in SMA patients. Screening of 100 control DNA samples revealed that four probands were carriers of only one SMN1 copy: the mean SMN1 dosage value was between 0.70 and 1.02, thus corresponding to the range typical for carriers of the SMN1 deletion. A further proband carried three SMN1 copies, the mean SMN1/factor VIII ratio was approximately 3.36. The mean SMN1 copy number of the other 95 controls ranged from 1.56–2.20, thereby confirming the reproducibility of the test as well as the distinct range that allows the exact determination of a SMN1 deletion carrier status.

The results in the carrier group confirmed the carrier status in all samples: the quantification range spans from 0.60 to 1.20. The distribution of SMN2 copy numbers in healthy controls: DNA samples of 38% of individuals gave an amplification pattern corresponding to one copy, the mean measured SMN2 copy number ranged from 0.80 to 1.20. Two copies were detected in 48% of samples (range from 1.60 to 2.32); three copies were observed in 5% (range from 2.72 to 2.92). In another nine cases, homozygous deletions of the SMN2 gene could be observed. To evaluate the reliability of the quantitative SMN2 approach in case of higher copy numbers, we additionally screened 40 SMA patients. The measured copy numbers in SMA patients with one to three SMN2 copies were in the same range as those observed in the control groups. Additionally, we identified 11 carriers of four copies (dosage range from 3.64–4.12).

6.4 PCR-RFLP (multiplex PCR and real-time PCR) [27]:

The 44 patients were classified according to the age of onset and the severity of the disease into three types: SMA type I, 20 patients (45%); SMA type II, 13 patients (30%); and SMA type III, 11 patients (25%). PCR-RFLP analysis showed homozygous deletion of exon 7 of SMN1 gene in all the 44 patients.
Detection of homozygous deletion of exon 5 of NAIP gene: multiplex PCR analysis of exons 5 and 13 of NAIP gene revealed homozygous deletion of exon 5 in 26 patients (45%, 26/44), 12 patients with SMA type I (60%, 12/20), 6 patients with SMA type II (46%, 6/13), and 11 patients with SMA type III (73%, 8/11).

SMN2 gene copy number analysis

Real-time PCR was performed to determine the SMN2 gene dosage, through the detection of CNVs of exon 7 of the SMN2 gene using specific primers and SYBR Green master mix (Qiagen). Cycle threshold (Ct) values of all patients with the three clinical SMA types were normalized according to the constructed standard curve. The copy number variations measured were 1 copy, 2 copies, 3 copies, or more than 3 copies.
The number of patients with different SMA types and the corresponding CNVs of exon 7 of the SMN2 gene. It was found that there was a significant difference between the three clinical groups (SMA type I, II, and III); p-value < 0.00001. In addition, 70% of type I-SMA patients had 1 copy of exon 7 of the SMN2 gene.
In type III-SMA patients, 82% of patients had more than 3 copies of exon 7 of the SMN2 gene. In type I SMA, no patient had more than 2 copies of exon 7 of the SMN2 gene. Patients with one copy of exon 7 of the SMN2 gene were only of type I-SMA. Patients with two copies of exon 7 of the SMN2 gene were either type I or II SMA, whereas patients with 3 copies or > 3 copies were of either type II or III SMA.
Combining the genotypes of SMN2 and NAIP genes, 8 genotypes were determined:
No significant difference was detected between D1 and N1 genotypes. A significant difference was found in patients with type II SMA ($p = 0.018$) for the genotype N2. Patients carrying two SMN2 copies of exon 7 and with no deletion in exon 5 of the NAIP gene (N2 genotype) were related significantly to type II SMA, whereas patients with the D2 genotype were likely to have type I SMA. Moreover, all patients with the D3 genotype had type II SMA; however, patients with the N3 genotype had type III SMA. It was found that the D4 genotype is related to type III SMA rather than the N4 genotype ($p = 0.01$).

6.5 Algorithm resolve the CNs of SMN1 and SMN2 using short-read GS data [22]:

6.5.1 Common CNVs involving the SMN1/2 loci:

Since existing PCR or NGS-based methods focus primarily on the c.840C>T site, this tool assumed a copy-number approach based on the sequencing data from the whole genes. The

investigated the read depth across the ~30-kb homologous region harboring SMN1 and SMN2 genes in 1kGP samples.
The depth profile indicates that this entire region can be deleted or duplicated in the population. The exact breakpoints of this CNV are expected to alter from sample to sample due to the extensive sequence homology that lies within and beyond this region and can only be resolved in high resolution with long read sequencing.
In addition to whole-gene CNVs, this tool also found a 6.3-kb partial gene deletion encompassing both exons 7 and 8, The sequences at the breakpoint are identical between SMN1 and SMN2, so this deletion takes place at either chr5: 70244114–70250420 in SMN1 or chr5: 69368689–69375000 in SMN2.
Among the samples that contain this deletion, this tool pointed out read pairs from samples where one spanned the breakpoint, and the mate spanned at least 2/3 of SMN-differentiating bases. Since both exons 7 and 8 are deleted, SMN2Δ7–8 most likely has little or no biological function. Therefore, SMN2Δ7–8 is an important variant that any SMN CN caller should consider.

6.5.2 SMN1/2 CN calls using differentiating bases:

This method called the CNs of SMN1 and SMN2 at the 16 base difference sites between SMN1 and SMN2 in samples, and compared the CN calls for each position with the CN calls at the splice variant.
There was a significant difference between the concordance of calls in the two populations. Excluding the first one, there were 13 sites that had high (>85%) CN concordance with the splice variant site, there were seven sites only that had high CN concordance with the splice variant site. The concordance values were lower than those in the second population.
This is compatible with within-gene variation at many of these positions, this tool selected the splice variant site and the seven positions that were highly coherent with the splice variant site in both African and non-African populations to make SMN1 and SMN2 copy number calls based on the consensus of CN calls at the selected sites.

6.5.3 Validation of the SMN copy-number caller:

After developing this method, its accuracy was tested against results from orthogonal methods. 73 samples were sequenced with known SMN1 and SMN2 CNs measured by digital PCR, 45 samples with known results measured by MLPA, and also compared our CN calls with MLPA.
The SMN1 CN calculated with this tool calls agreed with digital PCR results in 99.8% of the samples, the SMN2 CN calls agreed in 99.7% of the samples, and SMN2Δ7–8 calls agreed in 100% of the samples.

6.6 Novel integrated digital PCR instrument [10]:

Copy numbers of SMN1, SMN2 and total SMN in 10 genomic DNA test samples from Coriell Cell Repositories were quantified by the SMA multiplex 4-color assay. All 10 Coriell DNA samples were run in triplicate on the Absolute Q dPCR System for assay validation and repeatability. The intra-assay variability—measured by %CV—between the SMN1, SMN2 and total SMN is low, revealing strong repeatability. The intraclass correlation coefficients (ICCs) were 0.993 (95% confidence interval (CI) 0.979–0.998) for the SMN1 gene assay, 0.995 (95% CI 0.985–0.999) for the SMN2 gene assay and 0.987 (95% CI 0.963–0.996) for the entire SMN assay. The SMN1 and SMN2 CNs measured with this multiplex assay are concordant with those provided by Coriell Cell Repositories, including one sample with 5 copies of SMN2. Besides, the 6 copies of total SMN for NA03814 correctly matched the corresponding sums of SMN1 and

SMN2 copy numbers. The total SMN CNs in all the samples except for NA11254 were equal to their corresponding sums of SMN1 and SMN2 copy numbers.
For SMN1 and SMN2 copy number validation of blinded clinical samples, Absolute Q and QuantStudio 3D array dPCR SMA assays were compared on a set of blinded samples derived from patient lines (where n = 15). This set of samples contained gDNAs from 13 SMA patients and two sampels of non-SMA controls. Using the SMN1 copy number assay, the Absolute Q dPCR correctly determinated all SMA samples with homozygous deletions of SMN1 (12 out of 15) within the cohort. The Absolute Q dPCR identified one of the patient-derived sample (MND10) as containing one copy of SMN1 gene and one copy of SMN2 gene. This sample was from an individual having a SMN1 missense mutation27. The Absolute Q dPCR SMN1, SMN2 and entire SMN copy number measurements of the blinded clinical samples were concordant with those results obtained by QuantStudio 3D array dPCR27,31, including a sample (MND12) with 5 copies of total SMN. Bland Altman analysis of the results from SMN1, SMN2 and total SMN probes revealed strong agreement between Absolute Q and QuantStudio 3D dPCR assays. For each assay, 93% of the test samples fell in the limits of agreement. Moreover, the total SMN copy numbers from 14 blinded clinical samples agree with the sum of the copy numbers from SMN1 and SMN2 assays. In a single sample (MND03), the total SMN copy number at exon 7 (4 copies) was more than the sum of the SMN1 exon 7 (0 copies) and SMN2 exon 7 (2 copies) copy numbers. Based on this result, this sample was hypothesized to have 2 copies of either SMN1 or SMN2 having a partial deletion of exon 7. The partial deletion of exon 7 in this sample was confirmed orthogonally by junctional PCR (data not shown) and by the GS23.
The optimized assay was first tested on the control DNA samples containing SMN1 only (NA17117), SMN2 only (NA23255), or both SMN1 and SMN2 (NA03815). Representative images of the dPCR results from the analysis software reveals strong resolution of the partitions. All of the SMN probes (SMN1 exon 7, SMN2 exon 7 and total SMN intron 1) were normalized against the reference gene (RPPH1) that contains 2 copies per genome. The multiplex SMA genotyping assay run on the Absolute Q reveals high specificity and sensitivity in classifying the control samples with SMN1:SMN2 copy numbers of 0:3 (NA23255), 3:0 (NA17117), and 1:1 (NA03815). For NA03815, the Absolute Q dPCR data revealed copy numbers for SMN1 exon 7, SMN2 exon 7, and total SMN intron 1 as 1.1, 1.1 and 2.2, respectively.

VII. FUTURE WORK

We conducted a lot of research on tools that diagnose muscular atrophy patients. And as we have explained, all those tools are employed by analyzing genomic data, specifically by detecting the copy number variations of the genes SMN1 and SMN2. Copy number variation analysis requires a lot of laboratory instruments and specialized machines and devices. These machines can handle the huge amount of data and analyze the human genomes efficiently. To be more realistic, it is very difficult to apply this analysis in practical life on real genomic data. It is not suitable for researchers in schools and universities in and students. It hinders interest in this dangerous disease and would keep us away from reaching many innovative solutions. In parallel, the emergence and spread of machine learning and deep learning technology is huge and elevating each day. Therefore, we plan to work on and conduct researches on how to implement a deep learning model that predicts neuromuscular diseases and classify their types, including spinal muscular dystrophy, using a Support Vector Machine Algorithm on microarray data, and this will be applicable and suitable for anyone who needs it. These algorithms don’t require huge device capabilities and will give chance to a lot of researchers to participate in the coming scientific discoveries.

VIII. CONCLUSION

This paper provides an overview of the uses of copy number variation where more than one tool is based on CNVs for detecting spinal muscle atrophy like Code-Seq Technology, TaqMan Technology, SMAca Tool, PCR-RFLP, algorithm resolves the CNs of SMA1 and SMN2 Using short read GS data and novel integrated digital PCR. We also explained how copy numbers can be used to detect and identify a muscular atrophy patient. We talked about the importance of genetic patterns, and their role in detecting muscular atrophy, including SMN1, SMN2, and NAIP genes can help more in the prognosis of the more severe SMA type rather than less severe types.
The study showed a preference for the combination of modifier genes as a prognostic genetic pattern for phenotype determination.
Determination of CNVs of exon 7 of SMN2 is of great importance in the placement of the effectiveness of the new therapy (nusinersen), as it depends on having at least one copy of exon 7 of the SMN2 gene.
Of all the six tools we reviewed, three tools significantly caught our attention; The Algorithm that resolves the CNs of SMN1 and SMN2 using short-read GS data, SMAca and CODE-SEQ.
The Algorithm that resolves the CNs of SMN1 and SMN2 using short-read GS data showed very high accuracy where the SMN1 CN calculated with this tool calls agreed with digital PCR results in 99.8% of the samples, the SMN2 CN calls agreed in 99.7% of the samples, and SMN2Δ7–8 calls agreed in 100% of the samples.
The SMAca tool shows an overall high accuracy (0.998) and an F1 score of 0.938.
The CODE-SEQ was unique as it didn't only succeed to measure the copy numbers of genes SMN1 and SMN2, but it also succeeded to detect silent carriers; where the homozygous deletions of SMN1 exon 7 were considered consistent with a diagnosis of SMA. Moreover, the heterozygous deletions of SMN1 exon 7 were considered consistent with a diagnosis of the individual being a heterozygous carrier for SMA.
The CODE-SEQ results also showed very high accuracy, where it correlated with the MLPA results for all 80 samples, providing 100% correlation.

REFERENCES

[1] Prior TW, Finanger E. Spinal Muscular Atrophy. 2000 Feb 24 [Updated 2016 Dec 22]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK1352/

[2] Oskoui M, Levy G, Garland CJ, Gray JM, O'Hagen J, De Vivo DC, Kaufmann P (2007) The changing natural history of spinal muscular atrophy type 1. Neurology. 69:1931– 1936

[3] Prior TW, Nagan N, Sugarman EA, Batish SD, Braastad C (2011 Jul) Technical standards and guidelines for spinal muscular atrophy testing. Genet Med. 13(7):686–694

[4] Wirth B (2000) 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

[5] Luo M, Liu L, Peter I, Zhu J, Scott SA, Zhao G, Eversley C, Kornreich R, Desnick RJ, Edelmann L (2014 Feb) An Ashkenazi Jewish SMN1 haplotype specific to duplication alleles improves pan-ethnic carrier screening for spinal muscular atrophy. Genet Med. 16(2):149–156

[6] Alías L, Bernal S, Calucho M, Martínez E, March F, Gallano P, Fuentes-Prior P, Abuli A, Serra-Juhe C, Tizzano EF (2018 Oct) Utility of two SMN1 variants to improve spinal muscular atrophy carrier diagnosis and genetic counselling. Eur J Hum Genet 26(10):1554– 1557

[7] Serra-Juhe C, Tizzano EF (2019 May) Perspectives in genetic counseling for spinal muscular atrophy in the new therapeutic era: early pre-symptomatic intervention and test in minors. Eur J Hum Genet 3

[8] Ceylan AC, Erdem HB, Şahin İ, Agarwal M. SMN1 gene copy number analysis for spinal muscular atrophy (SMA) in a Turkish cohort by CODE-SEQ technology, an integrated solution for detection of SMN1 and SMN2 copy numbers and the "2+0" genotype. Neurol Sci. 2020 Sep;41(9):2575-2584. doi: 10.1007/s10072-020-04365-x. Epub 2020 Apr 6. PMID: 32249332.

[9] Hassan, H.A., Zaki, M.S., Issa, M.Y. et al. Genetic pattern of SMN1, SMN2, and NAIP genes in prognosis of SMA patients. Egypt J Med Hum Genet 21, 4 (2020). https://doi.org/10.1186/s43042-019-0044-z

[10] Jiang, L., Lin, R., Gallagher, S. et al. Development and validation of a 4-color multiplexing spinal muscular atrophy (SMA) genotyping assay on a novel integrated digital PCR instrument. Sci Rep 10, 19892 (2020). https://doi.org/10.1038/s41598-020-76893-7.

[11] Anhuf, D., Eggermann, T., Rudnik-Schöneborn, S. and Zerres, K. (2003), Determination of SMN1 and SMN2 copy number using TaqMan™ technology. Hum. Mutat., 22: 74-78. https://doi.org/10.1002/humu.10221

[12] Daniel Lopez-Lopez,Carlos Loucera,Rosario Carmona,Virginia Aquino,Josefa Salgado,Sara Pasalodos,María Miranda,Ángel Alonso,Joaquín Dopazo
https://doi.org/10.1002/humu.24120

[13] Robert J. Sicko, Paul A. Romitti, Marilyn L. Browne, Lawrence C. Brody, Colleen F. Stevens,
James L. Mills, Michele Caggana and Denise M. Kay, Rare Variants in RPPH1 Real-Time Quantitative PCR Control Assay Binding Sites Result in Incorrect Copy Number Calls, The Journal of Molecular Diagnostics, 10.1016/j.jmoldx.2021.09.007, 24, 1, (33-40), (2022).

[14] Ogino, S. & Wilson, R. B. Genetic testing and risk assessment for spinal muscular atrophy (SMA). Human Genetics (2002). doi:10.1007/s00439-002-0828-x https://pubmed.ncbi.nlm.nih.gov/12436240/

[15] Lefebvre, S. et al. Identification and characterization of a spinal muscular atrophy-determining gene. Cell (1995). doi:10.1016/0092-8674(95)90460-3 https://pubmed.ncbi.nlm.nih.gov/7813012/

[16] Prior TW. Perspectives and diagnostic considerations in spinal muscular atrophy. Genet Med.2010; 12:145–152.

[17] Ogino S, Leonard DGB, Rennert H, Ewens WJ, Wilson RB. Genetic risk assessment in carrier testing for spinal muscular atrophy. Am J Med Genet. 2002; 110:301–307

[18] Hendrickson BC, Donohoe C, Akmaev VR, et al. Differences in SMN1 allele frequencies among ethnic groups within North America. J Med Genet. 2009; 46:641–644.

[19] Sugarman EA, Nagan N, Zhu H, et al. Pan-ethnic carrier screening and prenatal diagnosis for spinal muscular atrophy: clinical laboratory analysis of >72 400 specimens. Eur J Hum Genet. 2012; 20:27–32

[20] Finkel RS, Chiriboga CA, Vajsar J, et al. Treatment of infantile-onset spinal muscular atrophy with nusinersen: a phase 2, open-label, dose-escalation study. Lancet. 2016; 388:3017–3026.

[21] Mendell JR, Al-Zaidy S, Shell R, et al. Single-dose gene-replacement therapy for spinal muscular atrophy. N Engl J Med. 2017;377: 1713–1722

[22] Chen, X., Sanchis-Juan, A., French, C.E. *et al.* Spinal muscular atrophy diagnosis and carrier screening from genome sequencing data. *Genet Med* **22**, 945–953 (2020). https://doi.org/10.1038/s41436-020-0754-0

[23] Sumin Zhao, Yaoshen Wang, Xiuqing Xin, Zhonghai Fang, Linlin Fan, Zhiyu Peng, Rui Han, Chaonan Shi, Yixiang Zhang, Chuang Fan, Jun Sun, Xuelian He, Next generation sequencing is a highly reliable method to analyze exon 7 deletion of survival motor neuron 1 (SMN1) gene, Scientific Reports, 10.1038/s41598-021-04325-1, **12**, 1, (2022).

[24] Ceylan, A.C., Erdem, H.B., Şahin, İ. *et al. SMN1* gene copy number analysis for spinal muscular atrophy (SMA) in a Turkish cohort by CODE-SEQ technology, an integrated solution for detection of *SMN1* and *SMN2* copy numbers and the "2+0" genotype. *Neurol Sci* **41**, 2575–2584 (2020).

[25] Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G (2002 Jun 15) Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res 30(12): e57

[26] 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

[27] Hassan, H.A., Zaki, M.S., Issa, M.Y. *et al.* Genetic pattern of SMN1, SM N2, and NAIP genes in prognosis of SMA patients. *Egypt J Med Hum Genet* **21**, 4 (2020).

[28] Lopez-Lopez D, Loucera C, Carmona R, Aquino V, Salgado J, Pasalodos S, Miranda M, Alonso Á, Dopazo J. SMN1 copy-number and sequence variant analysis from next-generation sequencing data. Hum Mutat. 2020 Dec;41(12):2073-2077. doi: 10.1002/humu.24120. Epub 2020 Oct 14. PMID: 33058415; PMCID: PMC7756735.

[29] Anhuf, Dirk, Thomas Eggermann, Sabine Rudnik-Schöneborn and Klaus Zerres. "Determination of SMN1 and SMN2 copy number using TaqMan™ technology." *Human Mutation* 22 (2003): n. pag.