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

Carolin Emad, Martina Romany ,Kerolos Mamdouh , Sherry Raafat, Abdelrhman Ahmed ,

Mary Elias , Mariam Nader .

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 I, type II ,type III and type IV) These types range from 6 months old to 35 years old.

It is classified according to the severity of the disease and the age at which symptoms appear. 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]. 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 occurs 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 paper 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

OBJECTIVES

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 [1].

We help researchers develop current tools and discover new tools and algorithms to serve muscular dystrophy patients all over the world, This will help in discovering a new therapy , as the treatment of this ailment is very rare and expensive

Our objectives is Increase awareness of this dangerous disease that is spreading every day, which can reduce the severity of the disease or cure it if it is discovered early 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, which have received FDA approval for the improvement of SMA symptoms.

Finally, our most important goals, which we always strive for, is to highlight its importance of genetic pattern, and its role in detecting diseases specially muscular atrophy.

MATERIALS & METHODS

1-Code-Seq Technology [2] 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.

2-TaqMan Technology [3] 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.

3-SMAca tool [4] 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.

4-PCR-RFLP (Polymerase Chain Reaction - Restriction Fragment Length Polymorphism)[5], 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.

5-resolve the CNs of SMA1 and SMN2 Using short-read GS data[6], 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 (GS) data by analyzing read depth and eight informative reference genome differences between SMN1/2.

6-the novel integrated digital PCR instrument[7], 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..

RESULTS

1- CODE-SEQ technology: The copy numbers in exon 7 of SMN1 gene matched with MLPA results in all 80 samples giving 100% correlation. The assay accurately detected the presence/ absence of SNPs associated with "2+0" genotype in the reference samples. None of the tested clinical samples had these SNPs.

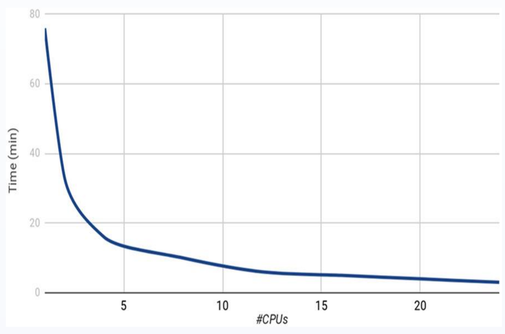
2- SMAca Tool: Results show an overall high accuracy (0.998) and an F1 score of 0.938. Seven samples (2.15%) were identified and successfully validated as putative SMA carriers. This tool also performed a large-scale validation of 1109 genomes from a diverse set of individuals.

3- TaqMan Technology: The reproducibility of the quantification approaches for exons 7 of the SMN1 and SMN2 genes were determined. Screening of 100 control DNA samples revealed that only four samples carried more than one copy of each gene. 40 SMA patients were additionally screened to evaluate the reliability of the quantitative approach in case of higher copy numbers.

4- PCR-RFLP (multiplex PCR and real-time PCR): The loss of both copies of exon 7 of the survival of motor neuron (SMN1) gene is the most common cause of SMA. This tool determined the copy number variations (CNVs) of SMN2 and NAIP genes in 44 SMA Egyptian patients.

5- resolve the CNs of SMN1 and SMN2 using short-read GS data:, 99.8% of our *SMN1* and 99.7% of *SMN2* CN calls agreed with orthogonal methods, with a recall of 100% for SMA and 97.8% for carriers, and a precision of 100% for both SMA and carriers.

6- Novel integrated digital PCR instrument: blinded cohort of 62 blood samples containing 21 SMA patients and 14 carriers, MS-CNV results were also highly concordant with MLPA. Both MS-CNV and MLPA quantified *SMN1* dosages without ambiguity.



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

CONCLUSION

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.
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

REFERENCES

[1] 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

[2] 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). https://doi.org/10.1007/s10072-020-04365-x

[3] 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

[4] 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.

[5] 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).

[6] 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

[7] 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.

ACKNOWLDGEMENTS

Dr. Tayseer Hassan, Eng :Mina Nagy .

The authors are grateful for the constructive comments and suggestions made by the reviewers