Accurately resolving copy number variation in highly homologous *SMN1* and *SMN2* genes using next-generation sequencing and rhPCR

H. Junnila, M. Valori, A. Korppoo, L. Koskinen, E. Salminen, K. Gall, T. Alastalo, S. Myllykangas, J. Koskenvuo, P. Salmenperä, and J. Sistonen
Blueprint Genetics, Biomedicum 1, Haartmaninkatu 8, Helsinki, 00290, Finland

Introduction

Spinal muscular atrophy (SMA) is a neuromuscular disorder characterized by progressive degeneration of spinal cord motor neurons, atrophy of skeletal muscles, and generalized weakness. The clinical phenotype ranges from a severe infantile form with a limited life expectancy to an adult-onset mild form of the disease. In the majority of cases, the disease is caused by the homozygous absence of the survival motor neuron 1 (*SMN1*) gene. The highly homologous survival motor neuron 2 (*SMN2*) gene differs from *SMN1* by a single coding nucleotide resulting in a splicing defect and reduced SMN2 function. *SMN2* copy number is a modifier of disease severity. Patients with a higher *SMN2* copy number usually present with a milder clinical phenotype. To address the clinical importance of accurate and efficient *SMN1* and *SMN2* copy number analysis, we developed a custom bioinformatic analysis based on next-generation sequencing (NGS) data combined with a novel RNase H2-dependent PCR (rhPCR) for confirmation analysis.

NGS-based Bioinformatic Analysis

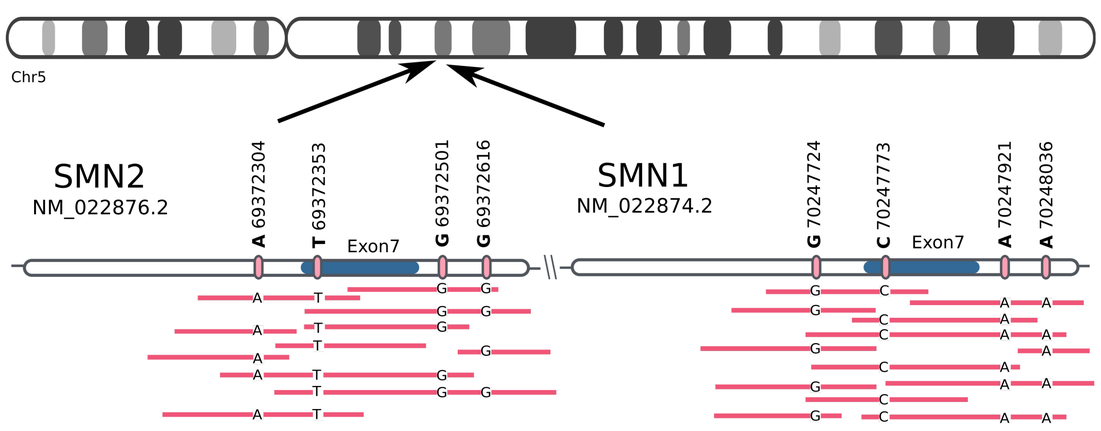

Figure 1. Whole exome sequencing was performed using the IDT xGEN Exome Research Panel with added custom clinical content and the Illumina NovaSeq platform. Bioinformatic *SMN1* and *SMN2* copy number analysis was based on calculating sequence read ratios at four nucleotide positions differing between the genes and inferring the copy numbers in relation to the normalized total *SMN* coverage modifying a previously published method [1]. Chromosomal positions are given in the Hg19 reference genome.

RNase H2-dependent PCR

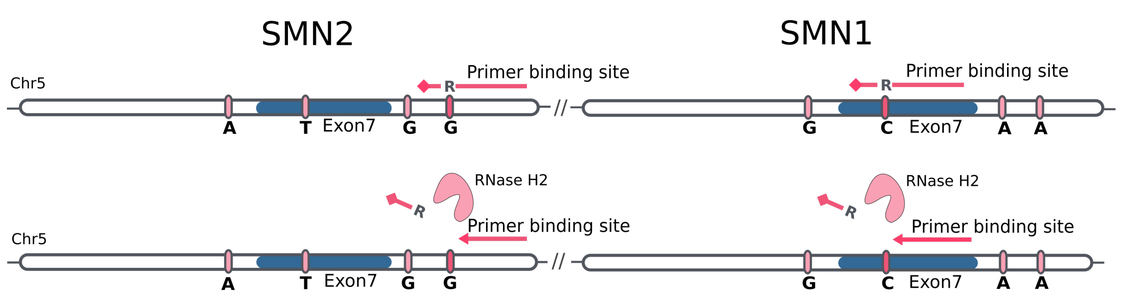

Figure 2. rhPCR utilizes blocked primers that are activated by RNase H2 enzyme after their hybridization to the target sequence [2]. RNA modification in the primer targeting a mismatched base allows improved discrimination between highly homologous sequences. Signal detection was based on 5'nuclease fluorescent probes and gene copy numbers were calculated using the ΔΔCt method similar to a standard quantitative PCR.

Results

Table 1. Analytical validation of bioinformatic and rhPCR methods by using 27 Coriell reference samples with known *SMN1/SMN2* copy numbers.

Copy number	NGS-based analysis identified correctly	rhPCR identified correctly
SMN1		
0	6/6	5/5
1	10/10	10/10
2	6/6	5/5
≥ 3	5/5	4/5
SMN2		
1	1/1	1/1
2	2/2	1/1
≥ 3	4/4	4/4
Accuracy	100.0%	96.8%
Sensitivity	100.0%	96.0%
Specificity	100.0%	100.0%

Table 2. Analysis of 89 clinical samples (55 blood, 6 saliva, 28 primary DNA) showed high concordance between the methods for *SMN1* copy numbers 0-2 and *SMN2* copy numbers 0-3 or higher. Performance of the methods was not affected by the sample material.

Copy number	Number of samples	Number of concordant calls between NGS and rhPCR	Concordance
SMN1			
0	2	2	100%
1	1	1	100%
2	70	69	99%
≥ 3	16	6	38%
SMN2			
0	9	9	100%
1	27	27	100%
2	48	48	100%
≥ 3	5	5	100%

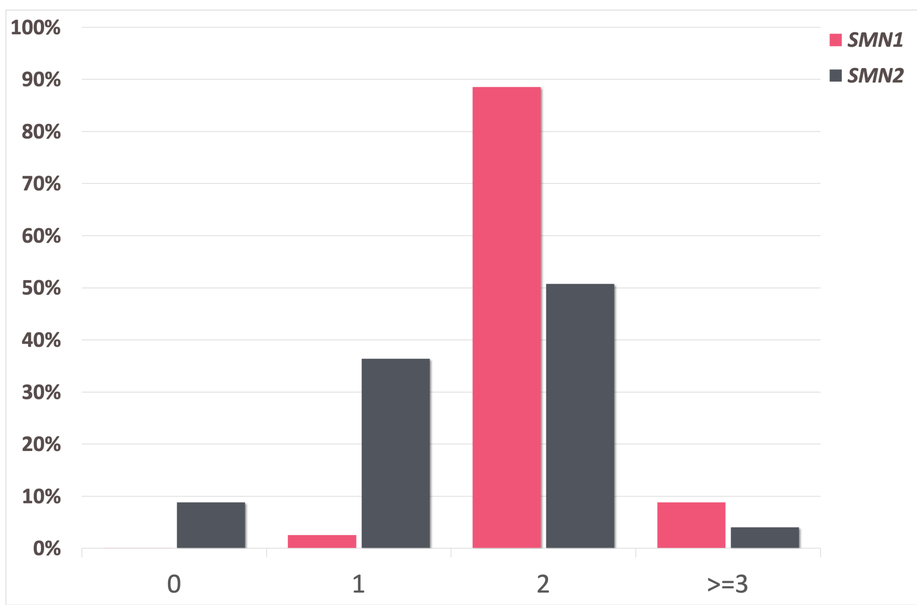

Figure 3. Frequencies of *SMN1* and *SMN2* copy numbers in 2196 de-identified samples submitted to Blueprint Genetics for genetic testing. Copy numbers were analyzed using whole exome sequencing data. Heterozygous *SMN1* deletion was observed in 2.6% of the samples, which is in agreement with frequencies reported earlier [3].

Summary

We have established an accurate and high-throughput approach to test for *SMN1* and *SMN2* copy numbers in blood and saliva samples enabling diagnostics of SMA and application of novel therapeutic strategies.

Blueprint Genetics

References:

1. Feng Y, Ge X, Meng L 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 Med* 2017;19:936-44
2. Dobosy JR, Rose SD, Beltz KR et al. RNase H-dependent PCR (rhPCR): improved specificity and single nucleotide polymorphism detection using blocked cleavable primers. *BMC Biotechnol* 2011;11:80.
3. Larson JL, Silver AJ, Chan D et al. Validation of a high resolution NGS method for detecting spinal muscular atrophy carriers among phase 3 participants in the 1000 Genomes Project. *BMC Med Genet* 2015;16:100.

Conflict of interest statement: All authors are employed by Blueprint Genetics.